Two-Dimensional Semantics

edited by

MANUEL GARCÍA-CARPINTERO
AND JOSEP MACIÀ

TWO-DIMENSIONAL SEMANTICS

This page intentionally left blank

Two-Dimensional Semantics

Edited by
MANUEL GARCÍA-CARPINTERO
and
JOSEP MACIÀ

CLARENDON PRESS · OXFORD

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trademark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2006

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data available
Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk
ISBN 0-19-927195-X 978-0-19-927195-5
ISBN 0-19-927202-6 (Pbk.) 978-0-19-927202-0 (Pbk.)
1 3 5 7 9 10 8 6 4 2

Contents

List of Contributors vii

1. Introduction 1
Manuel García-Carpintero and Josep Macià

2. Pragmatic Analyses of Anaphoric Pronouns: Do Things Look Better in 2-D? 22
Richard Breheny

3. Bad Intensions 38
Alex Byrne and James Pryor

4. The Foundations of Two-Dimensional Semantics 55
David J. Chalmers

5. Reference, Contingency, and the Two-Dimensional Framework 141
Martin Davies

6. Comment on 'Two Notions of Necessity' 176
Gareth Evans

7. Two-Dimensionalism: A Neo-Fregean Interpretation 181
Manuel García-Carpintero

8. Phenomenal Belief, Phenomenal Concepts, and Phenomenal Properties in a Two-Dimensional Framework 205
Martine Nida-Rümelin

9. Rationalism, Morality, and Two Dimensions 220
Christopher Peacocke

10. Indexical Concepts and Compositionality 249
François Recanati

11. Keeping Track of Objects in Conversation 258
Cara Spencer

12. Kripke, the Necessary Aposteriori, and the Two-Dimensionalist Heresy 272
Scott Soames

13. Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics 293
Robert Stalnaker

14. Two-Dimensionalism and Kripkean A Posteriori Necessity 310
Kai-Yee Wong
15. No Fool's Cold: Notes on Illusions of Possibility 327
Stephen Yablo
Index 347

List of Contributors

Richard Breheny, Lecturer in Linguistics, University College London

Alex Byrne, Professor of Philosophy, Massachusetts Institute of Technology

David J. Chalmers, Professor of Philosophy and ARC Federation Fellow, Research School of Social Sciences, Australian National University

Martin Davies, Wilde Professor of Mental Philosophy, and Fellow of Corpus Christi College, Oxford

Gareth Evans was Wilde Reader in Mental Philosophy, and Fellow of University College, Oxford

Manuel García-Carpintero, Professor of Philosophy, LOGOS-Logic, Language and Cognition Research Group, University of Barcelona

Josep Macià, Professor of Philosophy, LOGOS-Logic, Language and Cognition Research Group, University of Barcelona

Martine Nida-Rümelin, Professor of Philosophy, University of Fribourg, Switzerland

Christopher Peacocke, Professor of Philosophy, Columbia University

James Pryor, Associate Professor of Philosophy, New York University

François Recanati, Directeur de recherche, Institut Jean-Nicod, CNRS, Paris

Scott Soames, Professor of Philosophy, University of Southern California

Cara Spencer, Associate Professor of Philosophy, Howard University

Robert C. Stalnaker, **Laurance S. Rockefeller** Professor of Philosophy, Massachusetts Institute of Technology

Kai-Yee Wong, Associate Professor in Philosophy, The Chinese University of Hong Kong

Stephen Yablo, Professor of Philosophy, Massachusetts Institute of Philosophy

This page intentionally left blank

1

Introduction

Manuel García-Carpintero and Josep Macià

1. Kripke's 2-D Intimations

Kripke's (1980) *Naming and Necessity* convinced many philosophers that referential expressions like indexicals and demonstratives, proper names and natural kind terms are *de jure* rigid designators—expressions that designate the same thing with respect to every possible world. This feature distinguishes them from other singular terms like definite descriptions, which might also behave *de facto* as rigid designators, but *de jure* are not so. Kripke was well aware that his proposals created a philosophical puzzle. His view about referential expressions and alethic modalities entails the existence of *modal illusions*: truths that are in fact necessary appear to be contingent. Paradigm cases are instances of the schema ***if n exists, n is F***, with a rigid designator in the place of 'n' and a predicate signifying a hidden essential property of its referent in the place of 'F'. For the sake of illustration, let us replace 'F' in the schema with 'is-identical-to-Hesperus' and 'n' with 'Phosphorus':

(1) If Phosphorus exists, Phosphorus is-identical-to-Hesperus

The existence of those modal illusions elicited by Kripke's compelling views about referential expressions and alethic modalities is puzzling if we consider another compelling view about the epistemology of modality: that we have a reasonably reliable access to possible worlds. Kripke puts this as the intuition that a possible world "isn't a distant country that we are . . . viewing through a telescope . . . 'Possible worlds' are *stipulated*, not *discovered* by powerful telescopes" (Kripke 1980, 44); "things aren't 'found out' about a counterfactual situation, they are stipulated" (*op. cit.*, 49).[1]

We would like to thank Óscar Cabaco, Jose Díez, Dan López de Sa, Manuel Pérez and David Pineda for discussions on the topics of this introduction. We also wish to thank Óscar Cabaco who helped us prepare the Index, and Peter Momtchiloff at Oxford University Press and David Chalmers for their support and advice. This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the Spanish Government's grants DGI BFF2002-10164 and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409, from DGI HUM2004-05609-C02-01, DGI BFF2003-08335-C03-03, DURSI, Generalitat de Catalunya, SGR01-0018, and a *Distinció de Recerca de la Generalitat, Investigadors Reconeguts* 2002–2008.

[1] See also the analogous remarks in Kripke's (1980) preface that possible worlds are "given" by descriptive stipulations, pp. 15–20.

This puzzle is not an outright paradox constituted by contradictory claims; that one has in general a reliable access to modal facts allows for mistaken modal impressions. However, Kripke's views suggest that modal illusions do not arise only in a few, systematically unrelated cases; on the contrary, a systematic and far-reaching pattern is predicted. To sustain modal reliabilism requires thus a philosophical account of the illusions consistent with it. Kripke is sensitive to this, and, in his characteristically nuanced, cautionary mood, he provides one: "Any necessary truth, whether *a priori* or *a posteriori*, could not have turned out otherwise. In the case of some necessary *a posteriori* truths, however, we can say that under appropriate qualitatively identical evidential situations, an appropriate corresponding qualitative statement might have been false" (Kripke 1980, 142). In cases like (1), something more specific can be said:

> In the case of identities, using two rigid designators, such as the Hesperus-Phosphorus case above, there is a simpler paradigm which is often usable to at least approximately the same effect. Let 'R_1' and 'R_2' be the two rigid designators which flank the identity sign. Then '$R_1 = R_2$' is necessary if true. The references of 'R_1' and 'R', respectively, may well be fixed by non-rigid designators 'D_1' and 'D_2', in the Hesperus-Phosphorus case these have the form 'the heavenly body in such-and-such position in the sky in the evening (morning)'. Then although '$R_1 = R_2$' is necessary, '$D_1 = D_2$' may well be contingent, and this is often what leads to the erroneous view that '$R_1 = R_2$' might have turned out otherwise. (Kripke 1980, 143–4)

What Kripke proposes here, cautiously, only as a possible model applying in some cases, is the blueprint for 2-D accounts; the central idea is that "an appropriate corresponding qualitative statement", different from the original, necessary one, which unlike this "might have been false", is somehow mixed up with it, thus engendering the illusion of its contingency. Kripke refrains from making general claims about the applicability of this model. Nevertheless, his influential arguments against mind-body identity later in the *Naming and Necessity* lectures depend essentially on the premiss that the model is the only available one that properly explains the facts at stake.

This core 2-D idea can also be invoked to deal with the other puzzling Kripkean category of the contingent a priori, although Kripke's indications about this application are less clear. As he also famously noted, if one stipulates that a designator N is to be used to refer to an object introduced by a description D that thus fixes its reference, one can be said to know thereby a priori "in some sense" (*op. cit.*, 63) the truth of the corresponding statement 'N is D if N exists'; (2) provides an example, corresponding to (1):

(2) Phosphorus is whatever appears as shining brightly in the east just before sunrise, if it exists.

To apply the model here we should have that, although what (2) says is a contingent proposition, there is "an appropriate corresponding qualitative statement" which expresses a necessary one. This would provide for the partial rescue that Kripke (1980, 63 fn.) envisages for the traditional view that everything a priori is necessary.

Kaplan (1989) suggested related ideas, for specific examples of the contingent a priori involving indexicals, like 'I am here now' or 'I am the utterer'. He invoked a distinction between two different semantic features of context-dependent expressions, indexicals like 'I', 'here' and 'now' in particular: a *character* that captures the standing

meaning of the expression, and a *content* that consists of their truth-conditional contribution in particular contexts. Given a particular context, sentences like 'I am here now' express a contingent content; however, they are "character-valid" in that expressions in them have characters such that they will always express truths when uttered in any context.

Finally, Kripke suggested that the availability of (what we are presenting as his blueprint for) the core 2-D explanation of the necessary a posteriori and the contingent a priori supplies an important role for conceptual analysis, compatible with the Aristotelian-essentialist view that there are *de re* necessities which can only be known through empirical research:

> Certain statements—and the identity statement is a paradigm of such a statement on my view—if true at all must be necessarily true. One does know *a priori*, by philosophical analysis, that *if* such an identity statement is true it is necessarily true (N&N, 109). All the cases of the necessary *a posteriori* advocated in the text have the special character attributed to mathematical statements: philosophical analysis tells us that they cannot be contingently true, so any empirical knowledge of their truth is automatically empirical knowledge that they are necessary. This characterization applies, in particular, to the cases of identity statements and of essence. It *may* give a clue to a general characterization of *a posteriori* knowledge of necessary truths (Kripke 1980, 159).

Other writers, including Chalmers (1996), Jackson (1998, chapters 1–3), and Peacocke (1999, chapter 4) in particular, have subsequently elaborated on this idea.

2. Earlier Developments

Suggestions such as Kripke's and Kaplan's described in the previous section were taken up and developed in technically systematic ways in the two most influential articles originating the 2-D tradition after Kripke's inaugurating considerations, Stalnaker's (1978) "Assertion" and Davies and Humberstone's (1980) "Two Notions of Necessity".

Davies and Humberstone (1980) proposed to deal with the Kripkean phenomena by elaborating on Evans' (1979) related work on descriptive names and a distinction between *superficial* and *deep* necessity. According to Evans, intuitively a sentence is superficially contingent just in case the corresponding function from possible worlds to truth values assigns falsehood to some world; it is deeply contingent if just by understanding it one does not get any guarantee that there is a verifying state of affairs. The sentence 'snow is white if and only if snow is actually white' is an example of deep necessity combined with superficial contingency. Evans argued that, if one stipulates that 'Julius' is a name for the inventor of the zip, 'Julius invented the zip, if anyone uniquely did' is another case of deep necessity and superficial contingency. Davies and Humberstone (1980) considered modal languages that included the operator 'actually', to which they added a new operator, 'fixedly', intending thereby to represent Evans' views. As they emphasized, however, their proposals have a limited character; they could deal with cases of the contingent a priori and the necessary a posteriori involving descriptive names, but not with cases involving ordinary names, regarding which they accept Kripke's antidescriptivist arguments.

In "Assertion", Stalnaker introduced the well-known two-dimensional matrices to represent what he calls *propositional concepts*. He assumes that propositions, the contents of assertions and beliefs, can be modelled by means of classes of possible worlds. Thus, if we consider just two possible states of the world, the following matrix may represent the proposition expressed by (2); *i* is the actual state of the world, and *j* an alternative state in which it is rather Mars that appears as shining brightly in the east just before sunrise, but otherwise is as close as possible to the actual state of the world:

A

	i	*j*
	T	F

Worlds *i* and *j* illustrate one way in which the truth-value of what we say depends on the facts. However, there is a different way "that the facts enter into the determination of the truth-value of what is expressed in an utterance: It is a matter of fact that an utterance has the content that it has" (Stalnaker 1978, 80). If the facts had been different, what one says might have been different. Given the astronomical facts as they are relative to *j*, if the authority who in *i* has stipulated how the reference of 'Phosphorus' is fixed keeps that stipulation in *j*, (2) expresses a different proposition, one about Mars; we can represent this second way that the truth-value of what is expressed is determined by the facts in a second row:

B

	i	*j*
i	T	F
j	F	T

Matrix *B* represents what Stalnaker calls a *propositional concept*, a function from possible worlds into propositions (in their turn, functions from possible worlds into truth-values), or, equivalently, a function from ordered pairs of possible worlds into truth-values. Now, propositional concepts contain another perspicuous proposition, in addition to those represented by the horizontal lines: "I will call it the *Diagonal Proposition* since it is the function from possible worlds into truth-values whose values are read along the diagonal of the matrix from upper left to lower right. In general, this is the proposition that is true at *i* for any *i* if and only if what is expressed in the utterance at *i* is true at *i*" (*op. cit.*, 81). Stalnaker (1978) claims that, in this framework, an operator on propositional concepts "which says that the diagonal proposition is necessary . . . can be understood as the *a priori truth operator*, observing the distinction emphasized in the work of Saul Kripke between a priori and necessary truth. An a priori truth is a statement that, while perhaps not expressing a necessary proposition, expresses a truth in every context. This will be the case if and only if the diagonal proposition is necessary, which is what the complex operator says" (*op. cit.*, 83). This is naturally seen as a technical elaboration of the idea suggested by Kripke, that contingent a priori truths are statements whose straightforward content is contingent, while that of a corresponding appropriate qualitative statement is necessary; in our

example, the important "qualitative element" determining the relevant diagonal proposition is, of course, that the reference fixing stipulation obtains in all possible worlds in their role as context—those in the vertical axis. We will come back to this crucial matter presently.

The contrasting Kripkean phenomenon of the necessary a posteriori is the most important of the two from the point of view of the motivation that we found in Kripke's work for the framework, namely, accounting for the modal illusions engendered by his views on *de re* modality and thereby buttressing those views. Stalnaker appealed to his account of assertion (which uses the notion of speaker presupposition) together with Gricean pragmatic notions to deal, within the two-dimensional framework, with this phenomenon of the necessary a posteriori: "The presuppositions of a speaker are the propositions whose truth he takes for granted as the background of the conversation. A proposition is presupposed if the speaker is disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well. Presuppositions are what is taken by the speaker to be *common ground* of the participants in the conversation, what is treated as their *common knowledge* or *mutual knowledge*" (*op. cit.*, 84). This gives rise to a *context set*, "the possible worlds compatible with what is presupposed", "the set of possible worlds recognized by the speaker to be 'live options' relevant to the conversation. A proposition is presupposed if and only if it is true in all of these possible worlds" (*op. cit.*, 84–5). Stalnaker goes on to characterize assertions thus: "the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed" (*op. cit.*, 86).

Thus if a context includes just the worlds a, b and c, and an assertion is made in it that determines the following propositional concept

C

	a	*b*	*c*
a	T	T	F
b	T	T	F
c	T	T	F

the effect of the assertion will be to eliminate from the context set (from what it is presupposed) those worlds incompatible with what has been asserted. This means that if the assertion is not disputed, after it the context set will contain only the worlds a and b.

Stalnaker suggests a Gricean explanation of the modal illusion concerning (1) in two-dimensional terms. The straightforward Kripkean content of (1) is a necessary proposition with the following partial matrix:

D

i	*j*
T	T

Given the previously indicated facts about *j*, astronomical and linguistic, a corresponding propositional concept would be partially represented by the following matrix:[2]

E

	i	*j*
i	T	T
j	F	F

But the propositional concept in E does not offer a clear indication on how to modify the context set. According to E, what the speakers should do is: if the actual world is *i*, then keep both *i* and *j* in the context set; if it is *j*, then eliminate both *i* and *j* from the context set. Since the speakers do not know which of *i* or *j* is actual, they do not know how to proceed on the basis of E, and so asserting (1) would have no significant effect on the context set. It is thus pragmatically sensible to assume that the content asserted is not one of the horizontal propositions, but some other content, and the diagonal proposition offers a natural candidate; once again, it is a good candidate to represent the content of the "appropriate corresponding qualitative statement" that Kripke mentions.[3] Stalnaker appears thus to provide a nice technical elaboration, using only the possible worlds framework and other resources presumed to be equally important for any adequate theoretical account of linguistic and mental intentionality, of the Kripkean suggestions about how to reconcile compelling views on the metaphysics and epistemology of modality. As he summarizes it in retrospect: "necessary a posteriori truths, and contingent a priori truths could be represented by propositional concepts where the modal status of the diagonal diverged from that of the horizontal propositions expressed in the actual world" (Stalnaker 1999, 14).

[2] This is not completely accurate. Given their role in the vertical axis in *A* and *B* (i.e., that of possible contexts), worlds *i* and *j* should be taken as "centred" around transworld counterparts of the relevant utterance of (2); while now, in *D* and *E*, we want them to be centred around transworld counterparts of the relevant utterance of (1). We ask the reader to overlook these nuances here and elsewhere in what follows.

[3] Notice that in the case where the proposition determined in each world in the context set (the vertical axis) is the same, this proposition coincides with the diagonal proposition (if we restrict ourselves to the worlds that appear in the vertical axis). So Stalnaker could have said that the basic effect of any utterance U on the context set is *always* determined just by the diagonal proposition. On the other hand, if we want a matrix to represent whether its corresponding utterance is true or false, and the actual world is not in the context set (as will often be the case), then we should either (i) take the horizontal axis to include worlds beyond those that are in the vertical axis and, in particular, to include the actual world, or (ii) consider an "extended" diagonal proposition: one that, in addition to being defined for each world in the context set, it is also defined for the actual world. Options (i) and (ii) do not yield the same predictions regarding the truth of an utterance when, say, there is a failure of presupposition in the use of a definite description or a complex demonstrative (for instance, when I say "that woman is 5 feet tall", using "that woman" intending to refer to a person in front of us who we are assuming to be a woman, though in fact he is a man; if the person is in fact 5 feet tall, arguably option (i) would predict my utterance to be true, whereas option (ii) would predict it to be without a truth value).

3. Worries about the Explanatory Power of the Framework

The most fundamental issue about two-dimensionalism that has been raised subsequently is whether the fact that the diagonal of a statement *s* is necessary in any way accounts for *s*'s apriority. Surprisingly, perhaps, Stalnaker himself has been one of the most outspoken sceptics among those who have raised the issue: "I would emphasize now more than I did . . . what this abstract apparatus does not explain: one should not conclude that any account has been given of the nature of a priori truth and knowledge" (1999, 14).

If we go back to Stalnaker's original characterization of diagonal propositions, we can appreciate the root of this scepticism. In a depiction of an ordinary proposition like *A* above, the horizontal row represents the content of an utterance, given the facts relevant for the interpretation of the utterance in the actual world where it has been made. In order to motivate the introduction of further horizontal lines in the depiction, as in *B* above, Stalnaker reminds us that it is a contingent matter that a given utterance has a certain content; if relevant facts had been different, the very same utterance might have had a different content,[4] and this is what the different rows represent.

Now, to illustrate how the framework purports to explain the Kripkean phenomena, we previously kept fixed a crucial aspect of those facts determining the different contents that the very same utterance might have had: namely, the reference-fixing description tied to 'Phosphorus'. Without this, we do not get a propositional concept for (2) featuring a necessary diagonal proposition. However, there is nothing in the characterization of a diagonal proposition as such requiring it. The diagonal proposition is just defined as "the proposition that is true at *i* for any *i* if and only if what is expressed in the utterance at *i* is true at *i*", that is, the proposition that is true at *i considered as actual*, as the actual world where the utterance takes place. This only appears to require that an utterance of the very same phonological or graphical type occurs in the relevant world, and expresses something. There might be a possible world *k* where 'Phosphorus' is stipulated to refer to the innermost planet in the Solar System, even if, otherwise, *k* is as close to the actual world as it could be, in particular *k* is such that in *k* Venus is still the brightest heavenly body seen in the morning. If we add a row and a column for *k* to *B*, we get:

F

	i	*j*	*k*
i	T	F	T
j	F	T	F
k	F	F	F

Now we realize that, if we allow worlds like *k* to determine propositional concepts, we will never get necessary diagonal propositions nor, therefore, a priori truths as

[4] The very same utterance, or rather an epistemic counterpart of it; see fn. 14 in Stalnaker's contribution and the text to which it belongs.

modelled in this framework. And adding them fits the interpretation that Stalnaker assumes for diagonal propositions, which he describes as *metasemantic*, as he clearly indicates in recent works:

> Even paradigm cases of truths knowable a priori (for example simple mathematical truths) will have contingent diagonals in some contexts, on the metasemantic account. Consider a context in which a person is uncertain about whether the intended meaning of a certain token of "7 + 5 = 12" is the usual one, or one that uses a base 8 notation, with the same numerals for one through seven. In some possible worlds compatible with the beliefs of this person, the token expresses the falsehood that seven plus five is ten, and so the diagonal will be contingent. More generally, any utterance, no matter how trivial the proposition that it in fact is used to express, might have been used to say something false, and a person might have misunderstood it to say something false. So the metasemantic interpretation yields no account or representation of a priori truth or knowledge, and does not depend on any notion of the a priori. ("Assertion revisited", this volume, 302–3.)

The distinction between semantic and metasemantic interpretations of diagonal propositions parallels another distinction he makes, among semantic theories, between *descriptive* and *foundational*: "A descriptive semantic theory is a theory that says what the semantics for the language is without saying what it is about the practice of using that language that explains why that semantics is the right one. A descriptive-semantic theory assigns *semantic values* to the expressions of the language, and explains how the semantic values of the complex expressions are a function of the semantic values of their parts." Foundational theories, in contrast, answer questions "about what the facts are that give expressions their semantic values, or more generally, about what makes it the case that the language spoken by a particular individual or community has a particular descriptive semantics" (1997, 535). The variations in content represented by the horizontal propositions in a propositional concept correspond in a semantic interpretation to differences determined by facts investigated by descriptive theories (these are the sort of differences among rows in a Kaplanian character); on the other hand, a metasemantic interpretation allows for differences among horizontal propositions that correspond both to facts studied by descriptive theories *and* to facts studied by foundational theories.

In some cases, including some involving indexicals, Stalnaker concedes that the semantic interpretation can support applications of the 2-D framework so as to provide the explanatory benefits advertised of it, chief among them that of accounting for the Kripkean phenomena. But given Millian views about proper names, which many philosophers including Stalnaker hold, only the metasemantic interpretation can be invoked in cases like (1) and (2) above. Thus Stalnaker is ultimately as sceptical as Davies and Humberstone about how far the explanatory power of the framework reaches. This should help to put in its proper dimension both Stalnaker's "increasingly strident expressions of scepticism about a priori truth and knowledge" (this volume, 303, fn. 12), together with the fact that he still presents the framework as providing the explanatory benefits that "contingent a priori truths correspond to propositional concepts whose horizontal propositions are contingent, but whose diagonal is necessary" (Stalnaker 1999, 14). Understood without proper nuances, we presume that he would deem remarks like this ill-considered; the main nuance being

that the "notion of a priori that this identification yields is at best a very local and context-dependent one" (this volume, 303, fn. 12).

Stalnaker's "7 + 5 = 12" example thus shows that it is not in general the case that if a sentence is intuitively a priori then it has a necessarily true diagonal. This example alone would be enough to show a difficulty for the 2D-framework *both* to elucidate the existence of necessary a posteriori statements *and* the existence of contingent a priori statements (as the elucidation of both kind of phenomena using the 2D-framework would require the truth of *both* directions of the biconditional: *s* is a priori true iff *u* has a necessarily true diagonal). On Stalnaker's view we have also examples showing the falsity of the conditional: if *s* has a necessarily true diagonal, then *s* is a priori true. Consider a case where the participants in a conversation take as common knowledge that, say, George Harrison died in 2001; then in each world in the context set it is the case that George Harrison died in 2001. Therefore an utterance of "George Harrison died in 2001" in that context will have a necessarily true diagonal, even if such a sentence is not intuitively a priori.

4. More Ambitious Developments and Varieties of Scepticism about the Framework

The main aspiration of the 2-D framework that we uncovered in the Kripkean original intimations was to reconcile the appealing Kripkean metaphysical and semantic views, which envisage substantive *de re* necessities, with equally intuitive views on the epistemology of modality, which in their turn require an explanation for the ensuing illusion that such substantive necessities are contingent. We have seen that the earlier writers who developed the framework, Davies and Humberstone (1980) and Stalnaker (1978), are in fact sceptical that it can sustain these aspirations. In recent work, Frank Jackson and David Chalmers embrace the most ambitious goals of—as Chalmers metaphorically puts it in his contribution to this volume—restoring a golden triangle between meaning, reason, and modality unravelled by Kripke. For that, two-dimensionalism must sustain a *Core Thesis*, that "for any sentence S, S is a priori iff S has a necessary 1-intension" (the 1-intension corresponds to Stalnaker's diagonal). Here apriority is understood as an idealized form of knowledge constitutively independent (in some philosophically pertinent sense) of experience.

Chalmers' contribution, "The Foundations of Two-Dimensional Semantics", provides a helpful detailed taxonomy of different interpretations of 2-D ideas. He classifies them into two main contrasting views, the *contextual* and the *epistemic* understanding of the framework: "the contextual understanding uses the first dimension to capture *context-dependence*. The epistemic understanding uses the first dimension to capture *epistemic dependence*." He persuasively argues that the contextual understanding, in any of the different versions he considers, cannot validate the Core Thesis. The main problem is posed by well-known examples ('utterance problems': "language exists", "someone is uttering", "I think" and so on) previously mentioned by Evans, Kaplan and others to similar effects. Chalmers

takes them to be a posteriori, no matter whether we consider their intensions with respect to worlds taken as counterfactual or actual; but contextual interpretations of the framework ascribe them necessary diagonals.

In the core of his paper, Chalmers goes on to develop the epistemic understanding, and to argue that it can validate the Core Thesis. Aiming at being as ecumenical as possible, so that philosophers of different persuasions can find some use for the framework, he explores different ways of conceiving the fundamental concepts at stake: the nature of the "worlds" along the vertical and horizontal axis, and the relation in which they stand to sentences when worlds are considered as actual, as opposed to when they are considered as counterfactual. In particular, he explores the consequences for the Core Thesis of taking the worlds in the first dimension to be just the worlds of the second dimension, "centred" around a relevant context (agent plus location, token expression), or rather constructing them from scratch by using epistemic notions as primitives. The second approach makes it easier to validate the Core Thesis, but will be unacceptable to philosophers who find obscure the relevant epistemic notions (apriority, in particular). The first will be more appealing to these philosophers, but it is unclear that it can validate the Core Thesis, in both directions. It is unclear that every statement with a necessary diagonal is a priori: mathematical and logical examples raise legitimate doubts on that score. Cases of empirically revisable a priori statements raise reasonable doubts about the other direction.

We have already had opportunity to mention Stalnaker's recent scepticism with respect to the more ambitious aspirations of the proponents of the 2-D framework. In his contribution to this volume, "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics", Stalnaker contrasts interpretations motivated by an internalist conception of intentionality with his own, motivated instead by contrasting externalist views. He counts Chalmers' epistemic understanding as manifestly internalist. He points out that this way of distinguishing among interpretations of the framework cuts across his own distinction between *semantic* and *metasemantic* interpretations of the framework, which he explores in the first part of the paper. As we said, indexicals offer the best case for the semantic interpretation; in such a case, in moving from world to world along the vertical axis, expressions keep features that descriptive theories of the language must ascribe them, and in that respect are constitutive of their meaning. Stalnaker thinks that this interpretation is not available for many expressions, including proper names and natural kind terms. In moving from world to world along the vertical axis, proper names and natural kind terms might change their referents, and thus the only semantic features that on Millian views (like his own) a correct descriptive theory of the language ascribes to them. For this case, only a contrasting metasemantic interpretation of the framework is reasonable, according to him. On the metasemantic interpretation, expressions only keep features that foundational semantic theories of languages ascribe to them. For proper names, a foundational theory will presumably appeal to the causal-historical relation between use and initial baptism that Kripke (1980) famously suggested. Thus, in moving from world to world along the vertical axis, a proper name will always be in some or other such a causal relation with its referent; but, to the extent that the latter changes, the expression altogether changes its meaning.

The second part of Stalnaker's paper raises problems for the epistemic interpretation, predicated on its internalist underpinning. Stalnaker begins by acknowledging that his discussion remains in a dispiriting abstract plane, blaming the indefiniteness of available characterization of conflicting theories of intentionality for that. He presents the main lines of David Lewis' version of what he calls "global descriptivism", which he takes to be the most thoroughly developed form of an internalist account. According to it, the content of utterances and beliefs is determined by whatever assignment yields the best fit between beliefs and world. Then he raises two objections, the *holism* problem and the *indirectness* problem. The first is the problem that, according to the view, different thinkers—or the same thinker at different times—ascribe different meanings to expressions. The second is the counterintuitive consequence that our access to meaning is very abstract.

Manuel García-Carpintero's contribution "Two-dimensionalism: a Neo-Fregean Interpretation" aims to show that, in spite of the reasons for scepticism that Stalnaker puts forward, there is a neo-Fregean interpretation of the 2-D-framework that makes it possible to elucidate appropriately the two kinds of phenomena presented by Kripke (and, so, that it is possible to use the 2-D-framework to attain the goal of reconciling Kripke's metaphysical and semantic views with his epistemological views). The aims of García-Carpintero's paper, therefore, include showing that under his neo-Fregean interpretation (i) there is a principled way of excluding worlds like *k* in propositional concept F above that would undermine the 2-D treatment of contingent a posteriori statements, (ii) the referent of the designators that appear in necessary a posteriori statements varies among different worlds considered as actual (that is, among different rows in the propositional concept).

Even if García-Carpintero wants ultimately to answer the challenge posited by Stalnaker regarding proper names, he first considers the case of complex demonstratives. The reason is that even if a complex demonstrative (like "that morning heavenly body") is the sort of expression whose contribution to the proposition expressed by an utterance of a sentence that contains it is just some individual, it is clear in this case that there is also some descriptive content semantically associated with the demonstrative (for example that it refers to a heavenly body which appears in the morning, and which is somewhat salient when the demonstrative is used). García-Carpintero takes this descriptive content to be a conventional implicature of any statement that contains the complex demonstrative. He then argues at length that even if the object the demonstrative refers to and the descriptive content conventionally implicated are both semantically associated with the complex demonstrative, only knowledge of the descriptive content is part of the knowledge that constitutes linguistic competence in the use of the complex demonstrative. Linguistic competence in the use of a complex demonstrative is compatible with less than complete knowledge regarding empirical facts that might influence what the referent of the demonstrative is. García-Carpintero then postulates that the relevant propositional concepts to be considered when trying to elucidate the Kripkean phenomena are the ones where the knowledge that constitutes linguistic competence is kept fixed along the vertical dimension, even though we allow for variations in other facts that might affect what the referent of a demonstrative is. This allows, in the case of complex demonstratives, to meet the aims

(i) and (ii) mentioned above. García-Carpintero then argues that if we adopt a neo-Fregean view of the semantics of proper names (according to which proper names, like complex demonstratives, contribute just an individual to the basic proposition expressed, even though they have also (metalinguistic) descriptive content associated with them), we can extend the results for complex demonstratives just mentioned also to proper names. The ensuing account of a priori knowledge is one that is contextual, even though it is compatible with maintaining a general distinction between a priori and a posteriori knowledge: this contextual notion of a priori allows for some statements maintaining their a priori status among all ordinary contexts (and these are the statements that are traditionally regarded as examples of a priori knowledge).

Stephen Yablo's contribution, ''No Fool's Cold: Notes on Illusions of Possibility'', raises problems for the 2-D account of modal illusions that may arise out of concerns similar to those of Stalnaker. He first argues that the approach cannot account for some illusions of possibility involving sentences explicitly about what is actually the case. Thus, for instance, it appears to be possible for gold to have had a different chemical makeup than it actually has. However, it is at least not obvious how this can be accounted for inside the 2-D framework, on any of the interpretations so far considered. For, at first sight at least, that appears to require a world such that, when considered as actual, gold has a different chemical makeup than it actually does; and this is, of course, absurd.

Then Yablo goes on to argue that this alleged problem for the 2-D explanation of illusions of possibility goes much deeper than examples involving explicit claims about actuality might suggest. For, he argues, the usual 2-D explanations of illusions of possibility in fact explain most of them too easily. 2-D explanations (of, say, the illusion that a given table, which is actually made of wood, may turn out to be made of ice; or that heat may turn out to be low mean molecular energy) fail to satisfy a ''psychoanalytic standard'' that Yablo finds Kripke at least willing to meet, and which he thinks is acceptable in any case. This is the criterion that subjects who fall under the illusion, when apprised of any purported explanation like the one that the 2-D account posits, should be prepared to accept that their intuition testifies at best to the alternative possibility that the alleged explanation provides. Now, there presumably are worlds *w* such that something that does not look at all in its perceptible properties like a wooden table (say, an icy-looking purely rectangular block of ice) produce in *w*-observers, constituted very differently from the way we are in the actual world, the same kind of perceptual evidence (say, phenomenally similar conscious experiences) as this wooden table produces in us. However, Yablo argues, the existence of such worlds does not satisfy the psychoanalytic standard vis-à-vis the explanation of the epistemic possibility that this table, actually made of wood, may turn out to be made of ice. But, for all they say, customary 2-D explanations may well rely merely on the existence of worlds such as *w*.

The psychoanalytic requirement thus reveals, according to Yablo, that the illusions of possibility that must be accounted for (the ones that are not so easy to explain) implicitly include a reference to actuality: what must be explained is how something made out of ice, *but with observable properties like those that this table actually has*, could produce evidence as if of a wooden table *in people constituted as we actually are*.

If Yablo is right in his earlier argument regarding the difficulties that 2-D accounts have in order to deal with illusions of possibility concerning claims explicitly about what is actually the case, this would show that the difficulty goes much deeper. He in fact concludes by arguing that, under widely accepted Kripkean assumptions concerning essence and metaphysical necessity, no account along 2-D lines could satisfy the psychoanalytic standard in many important cases—including that of the constitution of heat by high molecular energy. If so, an altogether alternative account of illusions of possibility should be looked for.

Martin Davies' contribution, "Reference, Contingency, and the Two-Dimensional Framework", begins by presenting the main features of the framework that Davies and Humberstone (1980) proposed to deal with the Kripkean phenomena, and its connections with Evans' (1979) work on descriptive names and with Evans' related distinction between *superficial* and *deep* necessity; Davies outlines how that modal framework represents Evans' claims and distinctions. The volume includes a letter that Evans had written to Davies, commenting on a draft version of Davies and Humberstone (1980). In it, Evans raises a number of objections, among them a concern that the operator 'actually' behaves as one of Kaplan's (1989) *monsters*, that is, as a context-shifting operator, unlike usual operators in natural language. Relatedly, Evans objects that, by resorting to the elucidation of his views in terms of the modal behaviour of 'fixed actually', Davies and Humberstone's proposal may run into the sort of utterance problems that, as we saw previously, Chalmers' contribution discusses at length. Davies suggests in his contribution that the problem might be avoided if, in explicating deep modalities, the truth of sentences relative to worlds considered as actual is taken as primitive, while the truth of utterances is defined in terms of it.

Davies goes on in effect to discuss how Chalmers' Core Thesis fares on his views. He starts by pointing out that there are *prima facie* reasons to doubt Evans' rejection of a priori but deeply contingent truths—that is, those with contingent diagonals, in an explicitly 2-D transposition of Davies and Humberstone's notions; and also that, in any case, there does not appear to be any independent argument for that rejection. In order to illuminate this and the contrasting question, whether deeply contingent truths are a priori, he introduces a Fregean framework modified in a way suggested by Graeme Forbes, so that it is states of affairs rather than truth-values that are the referents of sentences. In this framework, he argues that deep-modal features of expressions depend on properties of entities at the level of reference, while epistemic features depend on properties of entities at the level of sense. There is then some reason to expect that apriority entails deep necessity, modulo the reservations previously expressed, but no reason to expect that the other direction also obtains—given that sense determines reference, but "there is no route back" from reference to sense.

In the final part of the paper, Davies discusses the specific problems posed by proper names. He argues that, in the particular case of a descriptive name *N* introduced by means of the description *the D*, the framework can be correctly invoked to establish that sentences of the form *N is D if there is a unique D* are only superficially contingent; they are deeply necessary, in consonance with their intuitive apriority. Similarly, if *E* is an essential property of *N*'s referent, the framework can be invoked

to establish that *N is E* is deeply contingent, although superficially necessary, and to account thereby for the illusion of contingency. However, he remains as sceptical as Davies and Humberstone (1980) regarding whether these points could be extended to cover all analogous cases involving ordinary proper names (using a *descriptive names strategy*). In order to justify his scepticism he appeals to variations of Kripke's (1980) epistemic, semantic and metaphysical arguments against more recent descriptivist theories of proper names. Davies concludes by examining his resulting disagreement with Evans' later view (also emerging in the letter included in this volume) that there is no semantic reason to classify descriptive names and ordinary names in different categories.

So far in this presentation, we have taken epistemic and metaphysical modalities as predicates of linguistic items. This is our way of solving the problem that Kripke (1980) deals with by resorting to a studiously ambiguous 'truths' to refer to bearers of those modalities. For, in that way, we can have contrasting modalities applied to the same bearer: one and the same item being both necessary and a posteriori, or the other way around. However, as Kai-Yee Wong points out in his contribution, "Two-Dimensionalism and Kripkean A Posteriori Necessity", this is misleading. For most philosophers would not take linguistic items, but propositions (in some understanding of them) as primary bearers of truth-values and their modal restrictions. Are there, then, propositions that are both necessary and a posteriori, or the other way around? This way of setting the question poses a problem for 2-D strategies, because the core idea is to ascribe metaphysical and epistemic modalities to different items; a sentence, say, is necessary and a posteriori by being associated both with a necessary content and with a different a posteriori content. To deal with this difficulty, Wong suggests that epistemic modalities are predicated of propositions, not absolutely, but relative to sentences expressing them. Although he does not go on to elaborate on what exactly this relativization amounts to, this view appears to be related to Martin Davies' suggestion, previously mentioned, to distinguish the state of affairs referred to by a sentence from the thought it expresses, and to take deep modal properties to be predicated primarily of the former but epistemic properties to be predicated of the latter.

Alex Byrne and Jim Pryor's contribution, "Bad Intensions", is also related to Davies', this time in the latter's scepticism regarding the applications of the framework intended to deal with illusions of possibility involving proper names and natural kind terms, and thus with the prospects of Chalmers' and Jackson's ambitious support for the Core Thesis. They distinguish three different roles that the association by a speaker of some properties with an expression (like a proper name or a natural kind term) could play: (i) an a priori role, in that the speaker might be able to know that the referent of the expression has the properties (if it exists), armed only with her understanding of the expression and a bit of a priori reflection; (ii) a *Fregean* role, if the association helps to solve relevant instances of the Fregean puzzle of the cognitive significance of true identity statements, and (iii) a *reference-fixing* role, if the association explains how the reference of the relevant uses of the expression is determined. They go on to present Chalmers' epistemic version of the 2-D framework, and to argue that it requires the existence of associations filling up the three roles.

In the core of the paper, they argue that Kripke's and Putnam's ignorance and error arguments refute this, as much as they refuted more naive versions of descriptivism, at least if the properties at stake are understood in the reductive, substantive form of the view to which Chalmers' usual examples appear to commit him. Byrne and Pryor take pains to insist that it does not help either to appeal to the merely implicit character of the association of expression and properties for the speaker, or to the non-linguistic character of the speaker's access to the relevant properties. In the final section, they discuss a weaker version of the view, a variety of the metalinguistic form of descriptivism allowing that the properties of the referent involve its relation to tokens of the expression itself; they argue that this also fails to fulfil the requirements of two-dimensionalists.

Scott Soames' contribution, "Kripke, the Necessary Aposteriori, and the Two-Dimensionalist Heresy", is part of a thorough critical examination of two-dimensional views. Here he contrasts the core 2-D proposals to deal with the Kripkean puzzles, as we saw already intimated by Kripke himself, with an alternative he finds more cogent; this alternative rejects what he takes to be a crucial tenet of different two-dimensionalist proposals, namely, that every way that, for all we know a priori, the world might be is a way that the world genuinely could be.

In the central part of the paper, Soames goes on to criticize what he takes to be the more coherently articulated version of two-dimensionalism that embodies this tenet; he calls it *strong ambitious two-dimensionalism*. He states the claim characterizing this doctrine that is mostly responsible for the objections he makes, in the following way (his thesis T5): ***It is a necessary truth that S*** is true with respect to a context C iff the secondary intension of S in C is true with respect to all world-states that are possible relative to C. By contrast, ***it is knowable apriori that S*** is true with respect to C iff in C, the primary intension of S is knowable apriori; ***x knows/believes that S*** is true of an individual i in C iff in C, i knows/believes the primary intension of S. Then he goes on to present four arguments against the view, based on examples in which modal operators interact with epistemic operators, causing trouble for the crucial two-dimensionalist tenet stated at the end of the previous paragraph. To give a flavour of these arguments, here is a simplified form of one of them. According to a strong two-dimensionalism including thesis T5, the following two statements (or corresponding ones) should have the same truth-values; however, while taking w to be a world such that Mars is the morning star, (3) might be true, (4) is false:

(3) John truly believes that Phosphorus/the actual morning star shines brightly in the east just before sunrise, but, had the world been in state w, Phosphorus/the actual morning star would not have shone brightly in the east just before sunrise and John would not have believed that Phosphorus/the actual morning star shone brightly in the east just before sunrise.

(4) John truly believes that Phosphorus/the actual morning star shines brightly in the east just before sunrise, but, had the world been in state w, Phosphorus/the actual morning star would not have shone brightly in the east just before sunrise and John would not have believed that the morning star shone brightly in the east just before sunrise.

Soames' paper also suggests further problems for other versions of two-dimensionalism, which have fewer commitments regarding the interaction of modal and epistemic operators with the primary and secondary intensions of the expressions on which they operate.

5. Some Applications of the Two-Dimensional Framework

This volume includes papers that cover three basic fields of applications of the two-dimensional semantic framework: the application to the study of the semantics and pragmatics of anaphora (Breheny's and Spencer's contributions), the study of concepts (Nida-Rümelin's and Recanati's contributions), and the study of morality (Peackocke's contribution). On the other hand, section 3.12 of Chalmers' contribution includes a brief summary of several other important applications as well as some references to relevant work.

The contributions of both Richard Breheny and Cara Spencer discuss ideas put forward in Stalnaker (1998). In this paper Stalnaker aimed to "describe the structure of discourse in a way that abstracts away from the details about the mechanisms and devices that particular languages provide for doing what is done in a discourse" (p. 97). To this aim he used his interpretation of the two-dimensional apparatus. As mentioned in Section 2 of this Introduction, Stalnaker understands context as common ground. The context set is the set of worlds compatible with what the participants in a conversation are assuming or *presupposing* at a certain point in the linguistic exchange. The context is constantly changing. The main point of an assertion is to reduce the set of possible worlds in the context set: if the assertion is not disputed those worlds not compatible with the asserted proposition will be eliminated from the context set. There is, though, another way in which an assertion changes the context: the very fact that the assertion is made and certain words are uttered will be common knowledge among speakers, and so this will be information that will be added to the context. Stalnaker (1998) considers how this framework can help clarify some uses of pronouns with indefinite antecedents, like the sequence of two sentences

(5) (i) I met a woman last night. (ii) She was from Portugal

(5i) is simply an existential claim. If accepted, it will be the case that in each world in the context set there is at least one woman that the speaker met last night (there might be more than one; it might be different women in different worlds). On the other hand "She" in (5ii) is a referential expression whose use presupposes that there is a woman uniquely available for reference. Stalnaker claims that the phenomenon of accommodation (that allows to incorporate into the context whatever non-controversial propositions are necessary to meet all the presuppositions that are required by the sentences used in a conversation) is at work in the interpretation of (5ii): the interpretation of (5ii) requires that there is a unique woman available for reference; by means of accommodation, after (5ii) is uttered a unique woman is made 'salient' or available for reference in each world in the context. In a felicitous use of (5), the speaker presumably will have a particular woman in mind when uttering

(5i). Therefore the obvious feature to appeal to in order to select a unique woman is that she is the unique woman that the speaker has in mind when uttering (5). It might be that two worlds in the context are exactly the same regarding the events that happened last night, say in both of them the speaker met exactly two women last night, but they are different regarding the additional information that the fact that the speaker has uttered (5ii) has brought into the context (through the mechanism of accommodation): in one of them it might be one of the two women that the speaker had in mind and, so, that is the woman who is available for reference, in the other world it might be the other woman that the speaker had in mind and who is, then, the one available for reference. Stalnaker (1998) shows how this basic idea can be used to clarify the uses of pronouns in several interesting examples.

Richard Breheny's contribution, "Pragmatic Analyses of Anaphoric Pronouns: Do Things Look Better in 2-D?", is concerned with the dependence between pronouns and indefinites, as in the discourse in (5) above. E-type approaches to the analysis of such discourses hold that the pronoun in (5ii) should be understood as if it were a definite description (for example, "the policewoman that I met last night"). Breheny distinguishes two kinds of E-type approaches: the *linguistic* approach and the *pragmatic approach*. According to the linguistic approach the definite description the pronoun goes proxy for is determined by a specific linguistic rule (one *very* simplified version of such rule would be: if a sentence contains an indefinite "a <noun phrase>", then a pronoun which is anaphoric to such indefinite (though outside its scope) is interpreted as the definite description "the <noun-phrase>"); according to the *pragmatic* E-type approach there is no linguistic rule that determines the definite description associated with the pronoun, and the process of recovering such description is purely pragmatic. In the first part of his paper, Breheny presents a series of data that are problematic for linguistic E-type approaches, though they can be accounted for on the pragmatic E-type approach.

In the second part of his paper, Breheny compares his pragmatic E-type approach with another pragmatic approach: a two-dimensional approach based on the ideas of Stalnaker (1998). Breheny points out that, even if Stalnaker (1998) does not explicitly mention this, it seems reasonable to assume that when interpreting (5ii) and given that the horizontal proposition determined at each world in the context might be different, the proposition that is expressed by (5ii) is obtained through diagonalization (in accordance with the Gricean mechanism we described at the end of Section 2 of this Introduction). Breheny, then, discusses some examples that would favour his pragmatic E-type approach over the pragmatic two-dimensional approach, such as the following:

(6) *(i) Last night I met two men. (ii) He was tall.

The use of "he" in (6ii) is infelicitous, though it seems that, following the two-dimensional ideas that allowed us to interpret the pronoun "she" in (5ii), the prediction would be that "he" in (6ii) is felicitous: (6i) ensures that there are at least two men the speaker met last night in each world in the context; then the utterance of (6ii) should bring about accommodation, and ensure that in each world one of the men the speaker met is made available for reference. On the other hand, the infelicity

of (6ii) does not present a problem for the pragmatic E-type account as, say, a use of the description "the person the speaker has in mind" that would be used to interpret the pronoun "he" would also be infelicitous. At the end of his paper, Breheny considers some possible moves within the general Stalnakerian framework that could be used to try to deal with the problematic examples that he has presented. He briefly argues, though, that none of these moves would, in the end, be successful, and he concludes that the pragmatic E-type approach should be preferred.

Cara Spencer's contribution, "Keeping Track of Objects in Conversation", addresses the question of what it is to keep track of what has been said at different points of a discourse about some particular objects under discussion. She considers the case of a conversation between speakers A and B in which at one point A utters "I got promoted"; then the conversation goes on, and later on B utters "you got promoted". Spencer considers the possibility of an eavesdropper that hears these two utterances but is unaware that they are part of the same conversation between the same speakers, and so misses that the speaker of one utterance is the addressee of the other. Spencer argues that the understanding that the eavesdropper has of the discourse is defective, and uses the Stalnakerian description of the discourse structure to explain exactly what the eavesdropper is missing. She argues that he is missing a presupposition (that is, a proposition that is part of the context), that can be identified as the diagonal of a certain special hypothetical identity statement that would contain A's utterance of "I", and B's utterance of "you". She discusses some metaphysical difficulties that arise in the application of the Stalnakerian framework, and contends that, if they can be solved, the Stalnakerian framework is better suited than a Russellian approach to appropriately describe the differences in the cognitive state of A, B, and the eavesdropper. In the final part of her paper, Spencer discusses a case where there is interaction between, on the one hand, the phenomena of keeping track of what is being said about a single object in the conversation and, on the other, the fact that at some point in a conversation it might be an open question which of two objects the conversation is about. Spencer describes the interaction that arises between her account of keeping track of objects in conversation and the Stalnakerian mechanism (described at the end of Section 2 of this Introduction) that determines that the proposition expressed is the diagonal proposition.

We now turn to the contributions that deal with applications of the two-dimensional framework to the study of concepts.

Martine Nida-Rümelin's contribution, "Phenomenal Belief, Phenomenal Concepts and Phenomenal Properties in a Two-Dimensional Framework", attempts to clarify the relationship between phenomenal properties (like the property of having a blue sensation) and phenomenal concepts of phenomenal properties (phenomenal concepts of phenomenal properties are in contrast with other concepts of phenomenal properties; for instance, with the concept of the property of having a blue sensation that a person who never had colour experiences might have acquired by talking to sighted people). In the first part of her paper, Nida-Rümelin elucidates and defends the thesis that to understand a phenomenal concept involves grasping the nature of the corresponding phenomenal property. She introduces a series of definitions and principles using the two-dimensional apparatus in order to clarify the notions that

appear in this thesis and also to clarify several additional notions that are, in turn, needed to explicate the notions that appear in the thesis. Just as a sample, we can mention that the paper presents two-dimensionalist characterizations of the notions of: *a concept C expressing a property P* (which consists in P being identical to the secondary intension of C), *grasping the nature of a property* (which is to have implicit knowledge of the secondary intension of some concept that expresses it), *understanding a concept* (which is to have implicit knowledge of the corresponding two-dimensional propositional function as a whole—that is, of the 'propositional concept' in Stalnaker's terms), *a concept being actuality independent* (which is for its corresponding two-dimensional function to meet the condition that: for any two worlds w and w*, the secondary intension we obtain if we consider w as actual, is identical with the secondary intension we obtain if we consider w* as actual).

Nida-Rümelin argues that phenomenal concepts (unlike, for instance, the concept of being water) are actuality independent. Given, therefore, the elucidation of *grasping a property* mentioned above, she can then finally justify the thesis that in the case of phenomenal concepts understanding the concept implies grasping the property that it expresses.

In the second part of her paper, Nida-Rümelin applies the notions she has introduced to the discussion of the possible replies that an identity theorist can offer to Frank Jackson's (1982) Mary argument. In particular she considers whether it is open to the identity theorist to claim that the property of having, for example, blue sensations can be grasped via a physical-functional concept that is available to Mary before leaving the room. Given various assumptions that Nida-Rümelin adopts, this possibility would arise only if it were possible for Mary after she leaves the room to find out that her concept of having blue sensations and her concept of having physical-functional property C are *necessarily* co-extensional. Nida-Rümelin shows that the sort of argument that one could use to justify the necessary co-extensionality of the concept of being water and the concept of being composed of H_2O cannot be used in this case, and that there does not seem to be any available alternative.

In his paper "Indexical Concepts and Compositionality" François Recanati characterizes a specific group of concepts, the indexical concepts, and argues that they are susceptible of a two-dimensional account similar to the one that can be provided for indexical expressions. Indexical expressions are associated with a rule that determines a content given a situation of utterance. Similarly, indexical concepts are also associated with a rule that determines the content of the concept on the basis of some specific contextual relation. The *kind* of contextual relation determines the type of the concept.

Recanati distinguishes different sorts of indexical concepts on the basis of the contextual relation on which they are based. These include, among others, *recognitional concepts* (such as the concept WATER, or the concepts that we have of each of our friends and other people we know), and *deferential concepts* (such as the concept of QUARK that even someone who does not really know what quarks are can still use to refer to quarks in thought). Recanati forcibly argues that these types of concepts are also indexical and that they can be then associated with both a character and a content. According to Recanati the very existence of these concepts

is contingent upon the existence of certain epistemic relations with the thinker's environment.

In the second part of his paper Recanati considers an argument by Jerry Fodor that would threaten Recanati's epistemic conception of indexical concepts. According to Fodor's argument, if we are to be able to give the usual compositionality account of productivity and systematicity, then nothing epistemic can be constitutive of concepts, because epistemic properties do not compose (the capacity to recognize water tanks in normal conditions does not depend on the capacity to recognize water in normal conditions). Recanati's response is based on drawing two distinctions: first, the distinction between compositionality of epistemic possession conditions (epcs) (if concepts C1 . . . Cn have epcs S1 . . . Sn, and they are the constituents of a complex concept D, then D has epistemic possession conditions and, furthermore, the epcs of D are a function of S1 . . . Sn) and what he calls *simple inheritance of epistemic possession conditions* (if S is an epc of a concept C, then S is an epc of any complex concept that has C as a constituent); and second, the distinction between compositionality of reference and compositionality of epistemic possession conditions. Complex concepts must respect simple inheritance but they do not need to meet compositionality of epcs. Only compositionality of reference and simple inheritance are necessary to account for productivity and systematicity.

We finally turn to Christopher Peacocke's contribution, "Rationalism, Morality, and Two Dimensions", which presents an application of two-dimensionalist ideas to the realm of moral discourse. In his contribution Peacocke defends that basic moral principles are a priori; more specifically, he defends what he calls the *Sharpened Thesis* that says that all moral principles that we know or that we are entitled to accept, are either *contentually* a priori or follow from both contentually a priori moral principles that we know and other non-moral propositions that we know. A *contentually* a priori proposition is, basically, one for which there is some a priori way of coming to know it that ensures that it is true whatever world is actual and regardless of whether it is judged or how it is judged. The judgement "I hereby judge that $13 \times 5 = 65$" (when it is reached not through introspection but by the thinker realizing on the basis of the concepts that it contains that it will be true whenever it is judged) is a priori but not contentually a priori—as it would not be true if worlds in which the thinker is not thinking were actual; on the other hand "$13 \times 5 = 65$" (reached by computation) is contentually a priori, as the computation method ensures that the content will be true whatever world is actual, and regardless of anything psychological.

Peacocke contends that theories that hold that there are mind-dependent properties that are constitutive of moral norms face the following challenge: to explain the apparent fact that moral principles are contentually a priori. Most specifically, mind-dependent theories face the challenge of explaining that moral principles are true in the actual world, whichever is the actual world. Using the two-dimensional apparatus, Peacocke distinguishes several ways of understanding this requirement. He focuses on two: (A) (diagonal reading) For any world w: P(w,w), and (B) For any world w: P($w_@$,w) (where "P$<w_1, w_2>$" denotes the truth value the proposition P would have when evaluated from the standpoint of the alleged

morality-generating attitudes of w_1, with respect to w_2). Peacocke argues that mind-dependent approaches of morality might be compatible with the reading (B), but not with the reading (A); the Sharpened Thesis, though, requires the truth of (A).

Peacocke offers a principle-based treatment of moral concepts: to possess a moral concept requires having an implicit conception whose content formulates, at least in part, what it is, constitutively, for something to fall under that moral concept. There are then ways of coming to know a basic moral principle that are guaranteed to be correct by the way in which the content of the relevant concepts is fixed. This is what accounts for the contentually a priori status of basic moral principles.

Peacocke stresses that his account avoids both a mind-dependent conception of moral truth and an epistemology that postulates a causal interaction with a 'moral realm'. This latter fact is what makes his approach to morality a *moderate* rationalist approach. Peacocke views the approach he defends in this paper as contributing to the development of a more general programme of moderate rationalism.

References

Chalmers, David (1996). *The Conscious Mind*, Oxford: Oxford University Press.

Davies, Martin and Humberstone, Lloyd (1980). "Two Notions of Necessity", *Philosophical Studies* 38, 1–30.

Evans, Gareth (1979). "Reference and Contingency", *The Monist* 62, 161–89. Also in his *Collected Papers*, Oxford: Clarendon Press 1985.

Jackson, Frank (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Kaplan, David (1989). "Demonstratives", in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 481–563.

Kripke, Saul (1980). *Naming and Necessity*, Cambridge, Mass.: Harvard University Press.

Lewis, David (1983). "Individuation by Acquaintance and by Stipulation", *Philosophical Review* 92, 3–32.

Peacocke, Christopher (1999). *Being Known*, Oxford: Oxford University Press.

Stalnaker, Robert (1978). "Assertion", in P. Cole (ed.) *Syntax and Semantics* 9, New York: Academic Press, 315–32; also included in Stalnaker (1999), from which we quote.

Stalnaker, Robert (1997). "Reference and Necessity", in D. Wright and B. Hale (eds.), *A Companion to the Philosophy of Language*, Oxford: Blackwell, 534–54.

Stalnaker, Robert (1998). "On the Representation of Context", *Journal of Logic, Language and Information* 7; also included in Stalnaker (1999), pp. 96–114, from which we quote.

Stalnaker, Robert (1999). *Context and Content*, Oxford: Oxford University Press.

2

Pragmatic Analyses of Anaphoric Pronouns: Do Things Look Better in 2-D?

Richard Breheny

1. Introduction

This paper is concerned with discourses of the form in (1), focusing on the anaphoric dependence between the pronouns in the second sentence and indefinites in the first.

(1) a. A man walked in the park. He whistled.
b. I predict that a woman will be nominated for President in 2008. Furthermore, I predict she will win.

The various analyses of such discourses can be classified as being either dynamic or not. Among the non-dynamic analyses, there are what will be called 'linguistic' and 'pragmatic'. A major concern of the first part of this paper will be to show that among non-dynamic approaches, there is good reason to think that some kind of pragmatic account might be on the right track. Though we aim to present some strong positive evidence for the pragmatic account which is also very problematic for dynamic approaches to meaning and discourse, we will not explicitly examine the question about the preferability of dynamic versus pragmatic accounts here.[1] Instead, we shall be interested in scrutinizing the virtues of two different pragmatic accounts. The first involves the E-type analysis of pronouns first proposed in Cooper (1979). The second, found in Stalnaker (1998), employs Stalnaker's two-dimensional framework. In the second part we will argue that Stalnaker's pragmatic account is, in an interesting way, too weak. We will consider and reject alternative, stronger proposals that could be made within the two-dimensional framework. We conclude, among other things, that only if pronouns can be considered to be E-type do we capture their definiteness.

Thanks to audiences at II Barcelona Workshop on Reference and S&B VI for useful feedback on this paper. Thanks particularly to Robert Stalnaker, Jason Stanley, Cara Spencer, James Prior, Manfred Kupfer and Emma Borg.

[1] The case against dynamic accounts of these kinds of examples is made in more detail in Breheny (2004). More general discussion of dynamic treatments of meaning and content is contained in Breheny (2003).

2. Non-dynamic Accounts

Among non-dynamic approaches to (1) we can distinguish between those which assume that the pronoun in the second sentence is bound by the indefinite in the first and those which do not. Among the latter, E-type accounts assume that the pronoun in the second sentence goes proxy for a definite description. E-type approaches in turn can be distinguished according to whether they treat the relation between indefinite and pronoun as mediated by linguistic rule.

2.1 Static binding

In terms of the traditional framework of semantic description, the only analysis according to which anaphoric pronouns such as in (1) are treated as variables has thus far been to assume that they are bound cross-sententially. The binding approach is motivated by the generally accepted intuition that (1) is understood according to the gloss in (2)a or according to the analysis in (2)b. This is the reading upon which Geach (1962) bases his proposals. It should be emphasized that it is not thought that there would necessarily be a uniqueness implication in such cases according to which just one man walked in the park:

(2) a. A man who walked in the park whistled.
b. $\exists x\,[man'(x) \wedge walked\_in\_the\_park'(x) \wedge whistled'(x)]$.

The idea that our understanding of such anaphoric relations is due to binding is further motivated by the apparent fact that pronouns which derive their interpretation from previous discourse are judged to be inappropriate where there is no proper antecedent. The infamous marble discourse, due to Barbara Partee, typically illustrates the point. (3)a below is judged infelicitous in spite of the fact that the pronoun in the final sentence is clearly meant to refer to the missing marble. Given that the antecedent sentences in (3)a and (3)b make available the same information, the contrast in acceptability suggests that there is something about the manner in which the information is presented which is the source of the unacceptability. Given the ancillary assumption that deictic pronouns are only properly used in the physical presence of their referent, there is a straightforward account of this given the binding approach: there is nothing to bind the pronoun in question, so it cannot receive an interpretation.

(3) a. I had ten marbles but dropped them. I found nine. ?It had rolled under the sofa.
b. I had ten marbles but dropped them. I found all but one of them. It had rolled under the sofa.

Without going into formal details, it is not difficult to be convinced that the binding approach suffers from a lack of generality. If the indefinite is treated as some kind of quantificational expression and the pronoun as a variable-like element, then one would assume that if cross-sentential binding can occur in the case of (1)a, it should occur in the case of other quantificational expressions. But this is not the case.

Consider that (4)a cannot be understood as (4)b, contrary to what we would expect if cross-sentential binding were a general phenomenon:

(4) a. Every boy left school early. ?He went to the beach.
b. Every boy left school early and went to the beach.

A more telling consideration involves certain quantificational antecedents as in (5). Here, the binding account predicts (5)a to be equivalent to (5)b and misses the fact that (5)a entails that just one boy left school early:

(5) a. Exactly one boy left school early. He went to the beach.
b. Exactly one boy left school early and went to the beach.

The E-type approach assumes that anaphoric relations as in (1) can result from the pronoun being understood as if it were a definite description. As Evans (1977) points out, the E-type approach correctly predicts our intuitions about (4) and (5), assuming that binding is only intra-sentential.

2.2 Linguistic versus pragmatic E-type approaches

Linguistic E-type approaches assume that the description in question is recovered by linguistic rule. With some proposals, the rule makes reference to the actual linguistic material in the antecedent sentence (see Evans 1977, Heim 1990). With others, the rule makes reference to the semantic interpretation of the antecedent expressions (see Neale 1990). Neale's proposal for a semantic rule is given in (6):

(6) "(P5) If *x* is a pronoun that is anaphoric on, but not c-commanded by '[Dx: Fx]' that occurs in an antecedent clause '[Dx: Fx](Gx)', then *x* is interpreted as the most 'impoverished' definite description directly recoverable from the antecedent clause that denotes everything that is both *F* and *G*." (Neale 1990, 182.)

The pragmatic alternative is to have no rule constraining which description the pronoun goes proxy for, leaving this matter to pragmatics—that is, to general principles of discourse plus particular facts about the context (Cooper 1979). Both approaches face problems.

2.3 Problems with uniqueness, contradictions and accessibility

The nub of the problem for the linguistic E-type approaches can be illustrated with the case of the missing uniqueness implications. The kind of syntactic rule for recovering the E-type interpretation which Evans proposes predicts that the discourse in (1) entails that just one man walked in the park. Similar problems arise for the semantic rules proposed by Neale. According to this rule, the pronoun in (1)a would be understood as in (7)b—contrary to intuition:

(7) a. He whistled.
b. [the$_{sing}$ x: man′(x) ∧ walked′(x)](whistled′(x))
c. *[the$_{sing}$ x: F(x)](G(x))* is true iff |F − G| = 0 & |F| = 1

The obvious diagnosis of the problem is that in these cases where there is no uniqueness implication, a speaker's referent of some sort is being introduced in the first part of the discourse to which reference is subsequently being made in the second part. So, one way to resolve this problem for linguistic E-type approaches might be to suppose that the indefinite in the first sentence is implicitly contextually restricted in such a way that this speaker's referent can be picked up in some way by the E-type description in the second. To illustrate how this proposal might work, suppose that sg_u expresses the property of being the individual which a speaker who makes an utterance u of an indefinite has 'in mind'. We can think of a speaker's referent as instantiating this property. What having an individual in mind amounts to in the general case is perhaps slightly problematic, but for the purposes of this paper we will assume that in factual utterances sg_u expresses the property of being the actual causal source of the intention underlying the speaker's utterance. Note, also, we assume that it is in the nature of sg_u that it be uniquely instantiated, if at all. The proposal would then be that the first sentence in (1)a is understood as in (8):

(8) $[\text{an x: man(x)} \wedge \text{sg}_u\text{(x)}](\text{walked\_in\_the\_park(x)})$

As such, subsequent E-type pronouns could be used to make reference to this speaker's referent. This would be possible on Neale's semantic approach to recovering the E-type interpretation, if not on Evans' syntactic approach.

While overcoming the uniqueness problem, this proposal is not viable since it makes the truth of the proposition expressed by the utterance of the first sentence dependent on how things stand with this speaker's referent regarding being a man and walking. It is well known that this is intuitively the incorrect analysis, for even if the speaker mistakes a woman for a man and thinks of that person as a man walking in the park at the time in question, then the proposition expressed by their utterance of the first sentence would still be true so long as there were men walking in the park at the time.

A related problem for the linguistic E-type approaches has to do with contradictions using pronouns. We note that a speaker (B in (9)) can coherently contradict another speaker (A) if they think the person they have in mind does not have the property used to describe them with an indefinite. However, on any kind of linguistic E-type approach, it is part of what is expressed by B that the individual being made reference to with the pronoun has this property which the rest of the utterance denies they have. So one should find B's utterance contradictory, if this kind of E-type account were correct.

(9) A: Last night I met a Cabinet minister.
B: She was not a Cabinet minister.

Of course we could suppose that when a speaker utters the first sentence in (1)a, some assumption about her grounds is communicated *implicitly* while the proposition expressed by the utterance does not depend for its truth on how things stand with the speaker's referent:

(10) What is said: $\exists x[\text{man}'(x) \wedge \text{walked\_in\_the\_park}'(x)]$

Implicitly communicated: $\exists x[sg_u(x) \wedge Bel(sp, man'(x) \wedge walked\_in\_the\_park'(x))]$

According to the pragmatic approach, such information could be exploited, resulting in an understanding of the second sentence in (1)a as in (11). According to the linguistic E-type approach, this is not possible:

(11) $\exists x[sg_u(x) \wedge whistled'(x)]$

As we will see presently, (11) is in fact a fair representation of our understanding of what the speaker expresses with the second sentence in (1)a. It is also clear that the cases where there is contradiction would not be problematic for the pragmatic approach as the description according to which we understand the pronoun would just involve sg_u.

While the pragmatic account does not suffer from the uniqueness problem and the related contradiction problem, it would be obliged to give an account for why the appropriate description can be recovered from implicit information in the case of (1)a but not in the marble discourse. This problem is quite severe since, if E-type pronouns are just like definite descriptions, then one would expect these pronouns to be acceptable in all cases where implicit information has to be exploited. That is, one would expect them to be understood via so-called bridging cross-reference—just as the descriptions are in (12):

(12) a. I had ten marbles but dropped them. I found nine. The missing one had rolled under the sofa.
b. Mary checked the picnic supplies and found that the beer was warm.

It is natural to think that there is something about the use of the indefinite that makes this speaker's referent suitably salient or the relevant information about the speaker's grounds suitably accessible. However, not very much has been offered in the way of defining what is or is not salient or accessible for a pronoun. Indeed, the severity of this problem has led many to think that the anaphoric relation between the indefinite and the pronoun is maintained in virtue of some kind of linguistic rule and not pragmatic inference.

To sum up, linguistic E-type accounts suffer from the uniqueness and contradictions problems but not the accessibility problem. By contrast, while pragmatic E-type approaches do not suffer from the uniqueness and contradictions problems, they are of questionable value unless a coherent story about salience or accessibility is provided. Indeed, this latter diagnosis applies to Stalnaker's *non*-E-type, pragmatic approach which we will review shortly. As the main purpose of this paper is to compare pragmatic approaches to such cases of anaphora, we will say little about the accessibility issue here. In Breheny (2004), a situation-theoretic approach to pragmatics and discourse is outlined according to which the appropriate distinction between types of implicit information can be made. This distinction does not involve any notion of (relative) salience. Within that framework it is possible to specify a presupposition for pronouns according to which the speaker's referent in (1)a is available for pronominal reference, but the missing marble in (3)a is not. We leave

this matter here, but, in order to motivate some interest in choosing between pragmatic accounts, the next section will review some positive evidence for the pragmatic accounts generally—evidence which is problematic for linguistic E-type and/or dynamic accounts.

3. More Evidence for a Pragmatic Approach

The parade examples which motivate linguistic accounts of pronominal anaphora tend to involve just one kind of language use: continuous, joined-up monologue. Turning away from such cases, it is a relatively straightforward matter to construct examples which parallel the marble discourse but where there is no infelicity—as in (13):

(13) When John came into the room, he found Mary holding a bag of marbles and staring intently at the floor. "What's up?" asked John. "I had ten marbles in this bag, but I dropped them," replied Mary, lifting up the rug. "How many have you found?" "Nine." "Bummer." Now both John and Mary began searching the nooks and crannies of the room. After half an hour's searching, John turned to Mary, "Do you think it could have rolled into the next room. . . ?"

Note, however, that it is not simply the dialogic nature of the above discourse that makes such antecedentless reference possible—as (14) demonstrates:

(14) John (manning a cake stall at a church fête, standing behind a lone cake): I baked six cakes and have already sold five of them. Mary (facing John with the lone cake between them, not looking down): *? It's my favourite kind. How much is it?

The generalization has to do with what the conversational participants are paying attention to. As pronouns contain no descriptive material, it would be unreasonable to refer to something with a pronoun that was not already in the focus of attention. Thus, the use of the pronoun pragmatically presupposes that what is being referred to is in the current focus of the audience's attention. In joined up, planned monologue, the speaker draws the audience's attention to one thing after another. In such cases, the speaker controls the focus of attention and it is up to her to ensure that the referents of pronouns are contained in what the audience is currently attending to. With indefinites, the linguistic meaning only characterizes a general type of situation. However, a speaker can indirectly indicate how they relate to a situation they are describing. Such indirect indications do not have a bearing on propositional content but they are part of the situation indicated. Such indirect indications can involve the speaker's grounds for what they say. See Breheny (2004) for a formal account of accessibility built on these ideas.

It is well known that where antecedent discourse contains utterances with quantificational expressions, say a sentence of the form *[[det[A]][B]]*, then pronominal reference can be made to members of the restrictor and intersective sets (that is $\{x: A(x)\}$ and $\{x: A(x) \wedge B(x)\}$) but not $\{x: A(x) \wedge \neg B(x)\}$. This is illustrated in (15), where the

second sentence cannot be construed as being about Clinton's non-supporters, despite the fact that world knowledge would push one to interpret it that way, if it was allowed.

(15) During the Lewinsky affair, most Democrats in Congress still publicly supported Clinton. Of course they represented more fundamentalist electorates.

This is really another accessibility fact which has either been passed over in dynamic treatments or dealt with in an ad hoc manner (cf. Kamp and Ryle 1993). What has not really been acknowledged before is that even in the case where a singular indefinite is used to introduce a speaker's referent, the members of the intersective and restrictor set are still available for reference as well. Consider (16)a, where the pronoun refers to riot policemen at the demonstration who cracked protesters' skulls; and (16)b, where reference is to riot policemen at the demonstration.

(16) a. At the Seattle demonstration, I saw a riot policeman crack a protester's skull for absolutely no reason. They should have been prosecuted for doing that.

b. At the Seattle demonstration, I saw a riot policeman crack a protester's skull for absolutely no reason. They all seemed to be under orders to club people at will.

This is entirely to be expected on the pragmatic account sketched above since it is assumed there that indefinites are just quantificational expressions and, as such, make these individuals available for pronominal reference—even where indefinites are used to introduce a speaker's referent implicitly. We can bring this point home by considering that the discourse in (17) is coherent but where reference is made both to members of one of these sets and to the speaker's referent.

(17) At the Seattle demonstration, I saw a riot policeman crack a protestor's skull. He just did it for no reason! They seemed to be under orders to club people at will./They should have been prosecuted for that.

This is all too expected on the pragmatic E-type account sketched above. This approach, however, does need to make some comment about cases, such as in (18), where pronominal anaphora is unacceptable with quantified antecedents.

(18) a. Every boy left school early. #He wanted to go swimming.
b. No boy left school early. #He was conscientious.

As Evans (1977) has observed, an E-type account correctly predicts the inappropriateness of (18)b as there is nothing for the description to quantify over. Regarding (18)a, Neale (1990) makes the observation that assuming the pronoun were E-type, an utterance of this discourse would be inappropriate as it would violate the manner maxim enjoining clarity. The reasoning behind this is that the use of the pronoun implies there is just one school boy. If this is the case, and given that a more appropriate form of words (e.g. 'The school boy' or 'The only school boy') was freely available, the utterance is confusing as it either sends mixed signals about the

number of boys or the pronoun has another, unknown referent. This kind of account gains independent support from cases where the maxim is flouted in order to make a joke—as in the case of (19) below, where Mandelson is a politician who is almost universally unpopular:

(19) Every Mandelson supporter was at the rally; but he was pretty lonely.

The pragmatic E-type account (and pragmatic accounts in general) says that where a discourse of the form, 'An F Gs. It Hs.' is understood as $\exists x[Fx \wedge Gx \wedge Hx]$ then this is due to the presence of a kind of implicature through which the speaker's referent is made available for the pronoun to refer to. There are a number of consequences that follow from this. First, we can ask what happens if, for some reason, there is no such implicit assumption. In that case, the E-type account predicts that the discourse would be understood to imply that just one F G-ed since this interpretive option is still available via what is explicitly expressed via the utterance of the antecedent sentence. To test this prediction, we need to consider cases where an indefinite is used but where we cannot reasonably assume that the speaker is implicitly communicating the relevant assumption about his grounds for using the indefinite. Two kinds of case come to mind. One is where the speaker just does not have specific grounds for an utterance involving an indefinite. The other is where, although the speaker may have specific grounds for what she says, this fact is not relevant or, more generally, not part of what the speaker can reasonably be assumed to be intending to communicate with the use of an indefinite. The latter kind of case can be illustrated with an adversarial discourse where, it is assumed, the speaker only gives away as much as is necessary. In (20) below, we have discussion between A and B about speed limits. We see that B's use of the indefinite does not imply he is thinking about any particular accident.

(20) A: If you'd ever witnessed a high-speed motorway accident, you wouldn't oppose the introduction of a speed limit.

B: I've spent half of my working life driving on motorways, so, in fact, I have witnessed a high-speed motorway accident. But I still think that one should be allowed to drive as fast as one wants.

If B were to follow up the general claim in this context with a statement using an anaphoric pronoun, then we would expect there to be a uniqueness implication (that he had witnessed just one such accident) since there is no implicature introducing a speaker's referent and so no option for the audience to understand the pronoun as referring to some particular accident the speaker has in mind. Considering (21), we find that this is the understanding we get:

(21) B: I've spent half of my working life driving on motorways, so, in fact, I have witnessed a high-speed motorway accident. It was fatal. But I still think that one should be allowed to drive as fast as one wants.

It is worth noting that one would assume that if someone spends a lot of time on motorways, then they would most likely have witnessed a good number of these accidents. This suggests that one is not free to interpret such discourses as if the

interpretive options were a matter of some kind of linguistic ambiguity (as van Rooy (2001) seems to suggest). It is the rhetorical properties of the preceding discourse which determine whether a speaker's referent is available. Further examples of this are already available in the literature. (22) is adapted from Geurts (1997). Here the first part of the discourse contradicts some stance of the audience. What kind of grounds the speaker has for this act is not all that relevant to this purpose and so it is unlikely that there is any implicature involving grounds. So, when a pronoun is used in the next assertion, we once again are bound to use only the general quantificational information as a resource for constructing an interpretation for the pronoun—resulting in a surprising uniqueness implication:

(22) It is ludicrous to pretend that there has never been an accident on this motorway. We both witnessed it, remember?

When predictions involving indefinites are made, we often understand the speaker to have no specific grounds. Stalnaker (1998) uses the prediction in (1)b (repeated below in (23)a) to make this basic point. To get the force of the example, we are supposed to be assuming that the speaker has no particular woman in mind in making the prediction. Intuitions may be sharper with the variant in (23)b:

(23) a. I predict that a woman will be nominated for President in 2004. I also predict she will win.
b. I predict a woman will finish in the top twenty in this year's marathon. But I predict she won't win it.

Another observation we can make about the pragmatic approach is that it makes the correct predictions with regard to what is explicit and what is implicit in discourses such as in (1)a. For these discourses, the prominent understanding is one which would be glossed, 'An F which Gs Hs'. According to the pragmatic E-type approach, the first sentence uttered expresses the general proposition involving the co-instantiation of F-ness and G-ness. While the claim about the second sentence is that it makes reference to the individual the speaker has in mind and says of it that it Hs. If that is the case, then the prominent understanding would have to be in some sense implicit. Can that be correct? To demonstrate that it is, consider the scenario in (24)a—adapted from Stalnaker (1998). Suppose also that, in reporting on the events, John utters (24)b:

(24) a. John is politically naive and is introduced by a practical joking host to a tabloid journalist as a cabinet minister and at the same time to a real cabinet minister as a journalist. In the ensuing (sincere) conversation, the real cabinet minister comes across as pro-Europe while the fake minister comes across as anti-Europe.
b. Last night I met a member of the Cabinet. He was anti-Europe.

While it would be appropriate for us to respond with (25)a below, we clearly could not respond with (25)b or c. So while it is clear that John is unwittingly misleading us

into thinking that he met a member of the Cabinet who was anti-Europe—nothing he actually says can be denied.

(25) a. He wasn't a member of the Cabinet.
b. You didn't meet a member of the Cabinet last night.
c. He wasn't anti-Europe.

It is important to note about (25)a and c that it is the speaker's intentions in introducing the referent into the discourse and not those of whoever uses the pronoun that determines the referent of subsequent pronouns. This point is illustrated in Stalnaker (1998) with (26). Here we find that A can coherently contradict what B says. If B were able to determine the referent of the pronoun, this would not have been possible:

(26) A: A man jumped off the cliff.
B: He didn't jump, he was pushed.
A: No not that guy, I know he was pushed. I was talking about another guy.

To sum up this survey of data, it seems that there is a quite robust generalisation that we understand "An F Gs. It Hs" according to Geach's gloss, $\exists x[Fx \wedge Gx \wedge Hx]$ only where the audience antecedently understands the speaker to be implicitly communicating something of her grounds—introducing an individual indirectly. Where this is not the case, there is a uniqueness implication. All of this is predicted by the pragmatic E-type account. It is also correctly predicted by the pragmatic E-type account that the 'Geachean' understanding is itself some kind of implicature. A fuller discussion of the pragmatics involved in these discourses and how other accounts, particularly dynamic ones, measure up when the full range of types of discourse are considered can be found in Breheny (2004). In this paper, we are interested in comparing pragmatic accounts. In the next section, we look closely at Stalnaker's two-dimensional pragmatic treatment which arguably handles the data discussed so far but with a more elegant and parsimonious analysis of pronouns.

4. Two-dimensional Pragmatic Accounts

In 'On the representation of context', Stalnaker (1998) sets out how a pragmatic account of (1)a and (1)b would go within his two-dimensional framework. Although Stalnaker does not explicitly work through his account assuming any particular analysis of pronouns, it seems clear that it is possible within the framework to treat pronouns simply as variable terms of direct reference and that this is the analysis which Stalnaker has in mind. As a simple, variable analysis may be perceived as more elegant and appealing, there may be motivation for favouring the two-dimensional framework for pragmatic analysis. However, we will see that, independently of the analysis of anaphoric pronouns, Stalnaker's account is interestingly too weak.

According to Stalnaker, discourse takes place against the backdrop of what the speaker assumes is commonly assumed. The set of possibilities consistent

with this 'speaker presupposition' constitutes the context set for an utterance. An assertive utterance can be thought of as an action directed primarily toward what is presupposed, reducing the set of possibilities in line with the content of the utterance. Not only the subject matter of a discourse can affect what is presupposed but also facts about the utterance itself. So, even if an assertion is rejected, facts about the attempt will affect the context set.

Standing assumptions among those presupposed will include Gricean-style discourse principles. Stalnaker (1978) includes a version of the principle that the hearer should be able to grasp which proposition the speaker expresses. In the 2-D framework, this amounts to the constraint that the speaker express the same proposition in each possibility in the context set. Note that, if we assume that the pronouns in (1)a and (1)b are both understood simply as variable terms of direct reference, it is clear that, on one level, we do not recover what proposition is actually expressed by the utterance of the second sentence in either case—as this would be a singular proposition depending on the same individual for its truth in each possible world in the context. As we have just seen, our understanding of the utterance of the second sentence is a descriptive proposition which depends for its truth on the individual which is the value of a function from contextual alternatives to individuals. In the case of (1)a, that function maps possibilities onto the individual the speaker has in mind in that possibility. In (1)b, it maps possibilities onto the unique woman nominee in that possibility. In both cases, our understanding is consistent with different individuals being the designatum of the pronoun in different alternatives in the context set. The disparity between what is literally said (on the variable analysis of pronouns) and the intuitive content of the utterance can be given a 2-D account as we can assume that we arrive at our understanding of these discourses through pragmatic processes which include what Stalnaker has elsewhere referred to as diagonalization.

In 'Assertion' Stalnaker proposes that a pragmatic process of diagonalization is available as a means of re-interpreting utterances which seem to violate some basic principles of discourse. In particular, he discusses how diagonalization would be employed as a repair strategy in the face of a flouting of the above-mentioned principle that the audience can grasp which proposition the speaker is expressing with her utterance. One case discussed involves A and B hearing the voice of an unseen person. A says, 'That is either Elizabeth Anscombe or Zsa Zsa Gabor'. Concerning the making of this utterance, there are three relevant alternative possibilities in the context set. There are possibilities where the demonstratum of the speaker's use of 'that' is Elizabeth Anscombe. There are those where it is Zsa Zsa Gabor. There are also those where it is neither. Given this, three different propositions would be being expressed by the speaker in different possibilities in the context, violating the above-mentioned principle. In this case, the speaker's intention seems to be to express what Perry (2001) would call the 'indexical proposition'—that the demonstratum is either EA or ZZG. As it happens, this proposition is the diagonal proposition of the two-dimensional, propositional concept for this utterance. In this case as always, the diagonal proposition is the proposition which is true in contextual alternative w when the proposition the speaker expresses in w is true in that alternative.

It is interesting to note that in the case of (1)a, the proposition expressed by the second statement in the discourse is not the indexical content. This can be verified by considering the contradiction examples above and others. As we will see, on Stalnaker's (1998) account of (1)a, the speaker expresses a contextually restricted diagonal proposition.

An outline of the reasoning behind our understanding of (1)a goes as follows:

(I) The assertion of the first sentence reduces the context set by eliminating worlds in which no man walked in the park. This effect is due to the conventional meaning of the first sentence uttered.

(II) After the first assertion is accepted, it is presupposed that in each live possibility there is an individual uniquely available for reference. This individual is that which the speaker had in mind in uttering the indefinite in the first sentence and this individual is a man who walked in the park.

(III) Then, in each possibility, when the second utterance takes place, the pronoun refers to that individual which is uniquely salient.

(IV) However, as the individual which is uniquely salient in each possibility is potentially different, the speaker would be failing to observe the above-mentioned principle, since potentially different propositions would be being expressed in different possibilities. So, the natural strategy at this stage is to diagonalize: assume that the speaker intends to convey not what is literally said, but the diagonal proposition.

(V) The resulting content of the utterance is equivalent to the proposition that the speaker's referent whistled.

Clearly, for this account to be acceptable, step II needs to be fleshed out a little. Stalnaker, in fact, does not go into all that much detail, assuming, as seems natural, that the details are largely self-evident. However, it is worthwhile considering what is supposed to be going on in these cases in somewhat more detail as we need to consider the role of other pragmatic strategies in addition to diagonalization and accommodation.

Supposing that pronouns are variable terms of direct reference, Stalnaker plausibly argues that such terms, when used, carry a pragmatic presupposition that there is an individual uniquely available for reference (see Stalnaker 1998). It is important to note that in the 2-D framework, if we assume that this presupposition is the only one which attaches to pronouns, it is not necessary that it is presupposed which individual the pronoun refers to in order to satisfy it (but only that in each possibility in the context an individual is uniquely available for reference).

So the presupposition for pronouns is quite weak. If it is accommodated without any further contextual reduction, we would get an understanding of the discourse in (1)a along the lines of, ''A man walked in the park. Some male whistled.'' So the question arises: How does context determine the descriptive understanding in question? The answer is that those possibilities where the individual available for reference is not the individual the speaker had in mind are ruled out on general pragmatic grounds. In this case and others, a principle of relevance or coherence (it amounts to the same here) does the job. This is so since in those possibilities where

the referent of the pronoun is not the speaker's referent, the speaker would—in the typical case—be failing to be relevant (or coherent).

We get an account of cases such as in (1)b, where we cannot presuppose the speaker has an individual in mind via similar reasoning. In these cases, the only way for the pragmatic presupposition associated with pronouns to be satisfied is if, in each possibility in the context set, there is a unique F which G-ed—hence the implication.

The elegance and parsimony of the linguistic analysis here is gained at the expense of slightly more complicated pragmatic reasoning. For note that this diagonalization account not only assumes that with discourses such as in (1) speakers do not say what they mean; in cases such as (1)b, they do not even mean what they say. Diagonalization comes into play here where the speaker flouts a kind of maxim of conversation. From Stalnaker's 2-D perspective, however, diagonalization is a natural strategy given that what is being updated includes information about the discourse as well as what the discourse is about. So, it is because the speaker is speaking, making reference, etc. in each possibility of the context, that, in cases like these, we make sense of the speaker's actions in the way we do in spite of the fact we may not be able to recover what they say.

There are, however, problems for this kind of account since it seems that context (including relevance/coherence principles) does not always do the work that it is meant to do. The problems arise with discourses where there is infelicity due to a kind of unresolvable ambiguity. At this stage it will make matters clearer if we distinguish between cases where the unacceptable indeterminacy arises because one cannot decide what the speaker intended as the source of relevance/coherence, and those where it is clear what the source of relevance would be (or how the segment is meant to cohere with preceding discourse). Infelicities of the former kind are illustrated in (27). Infelicities where the source of relevance is clear are illustrated in (28):

(27) a. Mary swore at Sue and she hit her.
b. A strong gust of wind blew the top of Mary's ice-cream onto Sue's dress. But she didn't notice.

(28) a. Mary's Hollywood dream was slowly turning into a nightmare of drugs and prostitution. She discussed her problems with Father Smith and Father Jones. ?* He wisely advised her to go back to her family's farm in Iowa and that's what she did.
b. ?* Two boys were playing cricket next door and he hit a shot which smashed my window.

The problem for the 2-D account is that one should be able to get a perfectly acceptable understanding of the pronouns in (28) as "one of them . . .". To see this, consider (28)a. Before the final sentence is asserted, in order to satisfy our expectations of relevance/coherence as well as the presupposition which attaches to the pronoun, all we need to do is to reduce the context set so that in some possibilities Father Jones is available for reference and Father Smith is in the others.

We could consider trying to strengthen the presupposition associated with pronouns so that it involves a notion of unique, maximal salience. However, it won't do to suppose that it is presupposed that the speaker is referring with a pronoun to the unique, maximally salient individual in the context, since it is clear that this is the wrong analysis. Consider the following examples:

(29) a. John can open Bill's safe. He knows the combination.
b. John can open Bill's safe. He'll have to change the combination.
c. Bill has a safe which John can open. He knows the combination.
d. Bill has a safe which John can open. He'll have to change the combination.

Experimental evidence suggests that, if anything, individuals referred to with the grammatical subject of an antecedent sentence are 'more salient' than those referred to by other arguments (see Gordon, Grosz, and Gilliom (1993)). But this does not preclude ambiguous pronouns from referring to less salient individuals in the context.

Moreover, even if the presupposition attaching to pronouns were that there is an individual uniquely available for reference and this individual is the maximally salient one, that still is not strong enough. After all, in (28)a, we could presuppose the one priest is maximally salient in some possibilities while the other is in the others. In so doing we would still reach a quite sensible understanding consistent with the principles or maxims of discourse.

It may not have gone unnoticed that the diagonalization strategy is not really necessary for cases like (1)a, as the interpretation of the pronoun could just as well be understood rigidly to be the individual at the end of the causal chain standing behind the speaker's intention. But, of course, this does not work in many other cases including ones such as in (1)b.

An analytical move to consider at this stage might involve pronouns being ambiguous between directly referential terms and descriptions. But this will only relieve the problem if diagonalization is not an option. Suppose we were to assume that pronouns could be descriptive. In that case, of course, all of the data discussed in this paper would be treated in a satisfactory way without the need for the extra pragmatic inference involving diagonalization, just as an E-type advocate argues. However, if in addition we were to assume that diagonalization were a general strategy available to participants anyway, then there does not seem any reason why it could not apply in cases like (28).

One can reasonably suppose that descriptive pronouns (like definite descriptions) carry an identifiability presupposition to the effect that the audience can recover what Perry (2001) calls an identifying condition satisfied by the designatum. But a moment's reflection would reveal that this is still too weak to rule out (28), given that we always have the option of diagonalization. In that case, although the pronoun's interpretation—a function from possibilities to individuals—could not be determined by the audience, a quite reasonable understanding of at least (28)a could be obtained by assuming that the description in question is either,

'the individual who is Father Jones' or 'the individual who is Father Smith', and diagonalizing.

5. Conclusion

There is quite strong evidence that pragmatic accounts of anaphoric relations between indefinites and pronouns are on the right track. In terms of the traditional framework for semantic description, the only analytical option available for a pragmatic account is to say that pronouns are E-type. Stalnaker has demonstrated that viewing discourse from the 2-D perspective opens up another analytical possibility given that what goes on in discourses like (1) makes diagonalization an obvious strategy to adopt. We have seen, however, that if diagonalization were a strategy we actually adopt, then we should be able to use discourses such as in (28) to convey existential information coherently and succinctly. Instead, it seems that pronouns do require the kind of strong identifiability which is incompatible with diagonalization being an openly available option. If that is so, and there is no other way to account for the unacceptability of (28), then we would have to say that pronouns must be E-type after all. Perhaps more seriously, it may be that diagonalization needs to be reconsidered, unless this kind of problem can be solved adequately.

References

Breheny, R. (2004). Indefinites and Anaphoric Dependence—A Case for Dynamic Semantics or Pragmatics? In M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*. Oxford: OUP, 455–83.

____ (2003). On the dynamic turn in the study of meaning and interpretation. In J. Peregrin (ed.), *Meaning in the Dynamic Turn*, Elsevier, 69–90.

Cooper, R. (1979). The Interpretation of Pronouns. In F. Heny and H. Schnelle (eds.), *Syntax and Semantics, Vol. 10: Selections from the Third Gröningen Round Table*. New York: Academic Press, 61–92.

Does, J. M. van der (1996). An E-type logic. In J. Seligman and D. Westerstahl (eds.), *Logic, Language and Computation Vol. 1*. Stanford, Ca.: CSLI, 555–70.

Evans, G. (1977). Pronouns, Quantifiers and Relative Clauses (I). *Canadian Journal of Philosophy* 7: 467–536.

Geach, P. (1962). *Reference and Generality*. Ithaca, Cornell University Press.

Geurts, B. (1997). Dynamic Semantics vs. DRT. *Zeitschrift für Sprachwissenschaft* 16: 209–26.

Gordon, P., B. Grosz and L. Gilliom (1993). Pronouns, Names and the Centring of Attention in Discourse. In *Cognitive Science* 17: 311–47.

Heim, I. (1990). E-type Pronouns and Donkey Anaphora. *Linguistics and Philosophy* 13: 137–77.

Kamp, H. and U. Ryle (1993). *From Discourse to Logic. Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.

Neale, S. (1990). *Descriptions*. Cambridge, Mass.: MIT Press.

Perry, J. (2001). *Reference and Reflexivity*. Stanford, Ca.: CSLI Publications.

van Rooy, R. (2001). Exhaustivity in dynamic semantics; referential and descriptive pronouns. *Linguistics and Philosophy* 24(5): 621–57.
Stalnaker, R. (1972). Pragmatics. In *Context and Contest*, 31–46.
____(1978). Assertion. In *Context and Contest*, 78–95.
____(1998). 'On the representation of context', *Journal of Logic, Language and Information* 7.
____(1999). *Context and Content*. Oxford: OUP.

3

Bad Intensions

Alex Byrne and James Pryor

1. Three Roles for Associated Properties

Let us say that a speaker *associates property P with word T* iff the speaker believes that the referent of T (if it exists) has P.[1] Here are three roles that associated properties might fill.

First, a speaker might be able to know that the referent of word T has certain properties (if it exists), armed only with her understanding of T and a bit of a priori reflection. If so, then let us say that those properties fill *the a priori role* (for word T). For instance, perhaps anyone who understands the word ***water*** is able to know, without appeal to any further a posteriori information, that ***water*** refers to the clear, drinkable natural kind whose instances are predominant in our oceans and lakes (if ***water*** refers at all—we will suppress this qualification from here on). Or, less controversially, perhaps anyone who understands ***water*** is able to know that ***water*** refers to a natural kind, or at least that it does not refer to an abstract object like a number. Or, almost *un*controversially, perhaps anyone who understands ***water*** is able to know that it refers to *water*. This last example shows that, plausibly, there will always be *some* property filling the a priori role for word T that its referent uniquely possesses—*being water*, in the case of ***water***. What is entirely unobvious is whether speakers have more interesting kinds of identifying knowledge about the referents of words: say, that ***water*** refers to the clear, drinkable natural kind predominant in our oceans and lakes. At first glance, such cases seem to be the exception, not the rule.

Frege's puzzle provides the second role for associated properties. As Frege pointed out in "On Sense and Reference," sentences like ***Bob Dylan is Robert Zimmerman***, unlike the sentence ***Bob Dylan is Bob Dylan***, "often contain very valuable extensions

Thanks to David Chalmers, Mike Nelson, Scott Soames, an audience in Barcelona, and two anonymous referees for helpful comments.

[1] Two points of clarification. First, the beliefs may be implicit, in the sense that the speaker would only judge that the referent of T (if it exists) has P upon ideal *a priori* reflection. More on this later. Second, for simplicity we will concentrate on singular terms, although the semantic theory ("two-dimensionalism") that is the topic of this paper is not so restricted. We will treat ***water*** as a singular term referring to a chemical kind. (We ignore predicative uses, as in ***O'Leary has some water in his basement***.)

of our knowledge.'' The ''cognitive significance'' (or ''informativeness'') of these sentences differ, and this is evidently because the cognitive significance of the name ***Bob Dylan*** differs from that of the coreferential name ***Robert Zimmerman***. To explain these differences in cognitive significance, many philosophers appeal to differences in the properties that speakers associate with the names ***Bob Dylan*** and ***Robert Zimmerman***. When the explanation of why T differs in cognitive significance from other coreferential words appeals to properties that the speaker associates with T, we will say that those properties fill *the Frege role* (for T).

Notice that properties that fill the a priori role need not fill the Frege role. The property *being Bob Dylan* (which is the same as the property *being Robert Zimmerman*), and arguably also the property *being sentient*, fill the a priori role for both ***Bob Dylan*** and ***Robert Zimmerman***. Since these properties are associated with both names, they cannot help explain the difference in cognitive significance between ***Bob Dylan is Bob Dylan*** and ***Bob Dylan is Robert Zimmerman***; accordingly they do not fill the Frege role.

Notice also that properties that fill the Frege role need not fill the a priori role. *Being the author of* Mr. Tambourine Man, for example, might fill the Frege role for ***Bob Dylan*** simply because it is a very well-known a posteriori fact that Dylan wrote *Mr. Tambourine Man*. Alternatively, *being the author of* Blow Ye Winds of Morning, might—at least in principle!—fill the Frege role for ***Bob Dylan***, for some speakers. But a speaker cannot *know* that the referent of ***Bob Dylan*** has this property, because Dylan *didn't* write *Blow Ye Winds of Morning*.

The question of *reference-fixing* provides the third and final role for associated properties. What makes it the case that the name ***Bob Dylan***, as we use it, refers to a certain person, namely Robert Zimmerman? (We may assume that this question has a non-trivial answer: it is not a brute fact that ***Bob Dylan*** refers to Robert Zimmerman.) The much-maligned *description theory of reference* gives one answer to this question. According to this theory, a word T (as used by a particular speaker) refers to an object *o* because the speaker gives a certain kind of reference-fixing authority to some properties $P_1, \ldots, P_n$. This makes T refer to whatever uniquely possesses $P_1, \ldots, P_n$—and that happens to be object *o*. When a speaker gives some of the properties she associates with T this kind of reference-fixing authority, we will say that those properties fill *the reference-fixing role* (for T).

Notice that it does not suffice, for some associated properties $P_1, \ldots, P_n$ to fill the reference-fixing role for T, that the referent of T is the unique possessor of $P_1, \ldots, P_n$. For properties to fill the reference-fixing role, the speaker has to (somehow) give them the special reference-fixing authority. (Of course, it is no easy matter to say exactly how a speaker might do this; for present purposes we can leave this tricky question aside.) Nor does it suffice, for $P_1, \ldots, P_n$ to fill the reference-fixing role for T, that the referent of T is the unique possessor of $P_1, \ldots, P_n$ *and* that $P_1, \ldots, P_n$ fill the a priori role for T. Properties can fill the a priori role for T without the speaker giving them reference-fixing authority. For example, the property *being water* fills the a priori role for ***water***, and water uniquely possesses it, but the speaker need not have fixed the reference of ***water*** to be whatever uniquely possesses this property. For present purposes, though, we can allow the converse. We can assume that speakers have some sort

of privileged access to the facts about what properties they have given reference-fixing authority to; and, hence, that any property that fills the reference-fixing role for T also fills the a priori role for T.

Notice that properties that fill the Frege role need not fill the reference-fixing role. We have already seen that a property that fills the Frege role need not be possessed by the referent (for example, *being the author of* Blow Ye Winds of Morning, in the case of ***Bob Dylan***). In addition, a property that fills the Frege role need not be uniquely identifying. (For example, perhaps *being a raspy-voiced singer* fills the Frege role for ***Bob Dylan***.)

Also notice that properties that fill the reference-fixing role need not fill the Frege role. Presumably someone could introduce ***Raspy*** as a nickname for Bob Dylan by giving the appropriate reference-fixing authority to the property *being Bob Dylan*. But, as we have seen, this property is associated with any name for Bob Dylan, and so does not fill the Frege role. We will mention another way of making the same point at the end of the paper.

So, with the one exception noted a few paragraphs back, there are no entailments (or, at any rate, no uncontroversial entailments) from filling one role to filling another. Moreover, for a given word T, although we may grant that *some* properties fill the a priori role for T, and that *some* (possibly distinct) properties fill the Frege role for T, it will often be controversial whether *any* properties fill the reference-fixing role for T.

Take ***water***, for example. Well-known arguments due to Kripke and Putnam appear to eliminate all the interesting candidates for filling the reference-fixing role for ***water***, for example *being the clear, drinkable natural kind predominant in our oceans and lakes*. All that remains are rather unexciting candidates like *being water*. And it is not at all obvious that even this property fills the reference-fixing role for ***water***. Of course, there will be *some* story to be told about why ***water*** has the referent it does; but the reference-fixing story we have been discussing is just one way this might be accomplished.

Given what we have said so far, it should seem rather implausible that a single set of associated properties could fill all three roles for a word. However, according to a sophisticated revival of the classical description theory—the semantic theory known as ***two-dimensionalism***—this implausible claim is actually *true*. For any word T, there are associated properties that simultaneously fill the a priori role, the Frege role, and the reference-fixing role. These properties are represented by a word's "primary" or "epistemic" intension: a certain function from possibilities to referents. Many proponents of two-dimensionalism take the theory to be something of a philosophical panacea, resolving a host of puzzles about language and thought—and posing a formidable challenge to physicalism into the bargain.

We think this enthusiasm is misplaced. Two-dimensionalism is incorrect basically for the reasons Kripke and Putnam gave thirty years ago, or so we will argue.

We will proceed as follows. Section 2 sets out the two-dimensionalists' central explanatory apparatus. We focus on David Chalmers' version of two-dimensionalism,

in particular his notion of "epistemic intensions."[2] Section 3 examines some considerations Chalmers gives for believing that words have epistemic intensions. We do not think that these considerations are persuasive. Section 4 briefly recapitulates part of the old, familiar case against the classical description theory, which can readily be adapted to apply to two-dimensionalism: Kripke's arguments from ignorance and error. Section 5 criticizes Chalmers' response to Kripke; and Section 6 examines a second response to Kripke, which we think also fails.

2. Epistemic Intensions

We now give a nuts-and-bolts summary of Chalmers' version of two-dimensionalism, making a number of simplifications for the sake of brevity.[3] In particular, we will ignore complications due to indexicals like ***I*** and ***now***.

An *epistemic possibility* is a hypothesis about how the actual world is, in respects that are left open by all one can know a priori. So, since the population of Barcelona is not an a priori matter, there is an epistemic possibility in which Barcelona has 1.1 million inhabitants, another in which it has 1.2 million, and so on. On the face of it, epistemic possibilities are distinct from the more common sort of metaphysical possibilities. Since it is not a priori that water is H_2O, there is an epistemic possibility in which water is, say, XYZ, and not H_2O, even though there is no such metaphysical possibility. In fact, Chalmers argues that the metaphysical possibilities and the epistemic possibilities are the *same* (minor qualifications aside); we will not be discussing this part of his view.

An *epistemically possible world* or *scenario* is a "maximal" epistemic possibility: an epistemic possibility E* that a priori implies all the other epistemic possibilities that are compossible with it.[4] (Henceforth, when we speak of "epistemic possibilities" we mean these "maximal" epistemic possibilities.)

The *epistemic intension* of a word T is a function from epistemic possibilities to objects that exist "in" or according to those epistemic possibilities. According to Chalmers, the value of T's epistemic intension at some epistemic possibility E may be

[2] Two-dimensionalism has also been defended recently by Frank Jackson (see especially his (1998a)). See Byrne (1999) for some discussion of Jackson's account. It has much in common with Chalmers' account, although there are some differences. For reasons of space, we cannot examine the differences here.

[3] For more careful expositions, see Chalmers (this volume), Stalnaker (2001), and Pryor (2003).

[4] In other words: E* does not leave any facts a priori open. For any epistemic possibility E, it is either (i) a priori that if E* is correct, then E is correct; or (ii) a priori that if E* is correct, then not-E is correct; or (an arguable qualification) (iii) a priori that if E* is correct, there is no determinate fact of the matter whether E is correct.

As will become clear shortly, the epistemic possibilities Chalmers officially defines his intensions over are specified in a very limited vocabulary (roughly: that of physics and phenomenology). Accordingly, it is entirely unobvious that these official epistemic possibilities are maximal in the sense just explained (not that Chalmers thinks otherwise).

determined by considering instances of the following schema (where ***t*** is replaced by the word T, and ***n*** is replaced by a singular term that appears in the specification of E):

> (Turns-Out) If E "turns out to be actual"—that is, if it correctly represents how the world really is—then *t* will turn out to be *n*.[5]

If (and only if) anyone who understands this conditional can know it to be true, perhaps after a bit of a priori reflection, then T's epistemic intension will be a function that maps E to the object *n*.[6] We will say that a speaker can *identify the referent of T in E* if and only if the speaker can know some instance of this schematic conditional to be true, in the way just described. In general, Chalmers supposes that for any word T, and any epistemic possibility E, anyone who understands T can identify its referent in E. As Chalmers and Jackson put it: an understanding of T by "a suitably rational subject bestows an ability to evaluate certain conditionals of the form $E \rightarrow C$, where E contains sufficient information about an epistemic possibility and where C is a statement using [T] and characterizing its extension, for arbitrary epistemic possibilities" (2001, 324, footnote omitted).[7]

Here are two examples Chalmers gives of identifying the referent of a word in an epistemic possibility:

> What about a term such as 'Hesperus'? ... Let scenario W_2 be one on which the brightest object visible in the evening is Jupiter, and where the brightest object visible in the morning is Neptune. For all we know a priori, W_2 is actual. If it turns out that W_2 is actual, then it will turn out that Hesperus is Jupiter. So when evaluated at W_2, the intension of 'Hesperus' returns Jupiter. If it turns out that A [the epistemically possible world that happens to describe the actual world correctly] is actual, then it will turn out that Hesperus is Venus. So when evaluated at A, the intension of 'Hesperus' returns Venus. (2002b, 145–6)

And similarly:

> Let W_3 be a 'Twin Earth' scenario, where the clear, drinkable liquid in the oceans and lakes is XYZ. For all we know a priori, W_3 is actual. If it turns out that W_3 is actual, then it will turn out that water is XYZ. So when evaluated at W_3, the intension of 'water' returns XYZ. If it turns out that A is actual, then it will turn out that water is H_2O. So when evaluated at A, the intension of 'water' returns H_2O. (2002b, 146)

(These reflections about what will turn out to be the case are supposed to be a priori.)

So, according to Chalmers, the epistemic intension of ***Hesperus*** differs from that of ***Phosphorus***, and the epistemic intension of ***water*** differs from that of ***H_2O***. He

[5] We assume that the conditional in this schema is the material conditional. We also assume that whenever E a priori implies that *n* exists, ***n*** appears in the specification of E. (Compare the "identifying descriptions" in Chalmers and Jackson (2001), 318.)

[6] On this formalization, *n* would always have to exist, because it is the value of a function that exists. Epistemic possibilities can, however, *say* that certain objects exist, which do not and indeed could not exist. This raises interesting questions about the ontology of epistemically possible objects. We cannot pursue those questions here, so we will assume for the sake of argument that they can be answered in a way that makes the notion of an epistemic intension coherent.

[7] The quotation actually concerns "concepts," rather than words, but clearly Chalmers and Jackson would allow the substitution. (See their footnote 7, p. 323.)

thinks that, in general, two words T_1 and T_2 have the same epistemic intension if and only if a speaker competent with these words can know that they are coreferential, armed only with her understanding of the words and a bit of a priori reflection. Since Chalmers takes synonyms to be words with the same epistemic intension, he also holds that if a speaker understands a pair of synonyms T_1 and T_2, she can know that they are coreferential. This claim is controversial, but we will not discuss it further here.

The apparatus of epistemic intensions is not supposed to be the whole semantic story, of course. *Two* "semantic dimensions" are required, because a word T also has a more familiar sort of intension: the function that takes a metaphysically possible world *w* to the referent of T at *w*. (That is, the function that delivers T's referent in possibilities taken to be *ways the world could, counterfactually, have been*, not *ways the world may be, for all one knows a priori*.) Since, necessarily, Hesperus *is* Phosphorus, and water *is* H_2O, the "metaphysical" or "counterfactual" intension of ***Hesperus*** is the *same* as that of ***Phosphorus***, and similarly for ***water*** and ***H_2O***.

We said that the epistemic intension of a word is determined by which instances of the schematic conditional like (Turns-Out) a speaker will be able to know a priori. What enables a speaker to know which of these conditionals are true, and which are false? We can think of matters like this. For any word T a speaker understands, there are some properties $P_1, \ldots, P_n$ that the speaker associates with T. More precisely, the speaker believes that the referent of T possesses $P_1, \ldots, P_n$ in the following sense: upon ideal a priori reflection, the speaker would judge that the referent of T possesses $P_1, \ldots, P_n$. These properties are such that the value of T's epistemic intension at epistemic possibility E is the object described by E as being the unique possessor of $P_1, \ldots, P_n$ (if there is such an object). According to Chalmers, any such properties will fill all three of the roles we mentioned earlier: the a priori role, the Frege role, and the reference-fixing role.

To illustrate these points, take ***water***. Going by the previous quotation, the associated properties are something like: *being clear*, *being drinkable*, *being in the oceans and lakes*. Since these properties fill the a priori role for ***water***, someone who understands ***water*** does not need any further a posteriori knowledge to know that the referent of ***water*** is clear, drinkable, and found in the oceans and lakes.

Since these properties fill the Frege role, the cognitive significance of a sentence like ***Water is H_2O*** derives from the fact that *being clear*, *being drinkable*, *being in the oceans and lakes* are associated with ***water***, and some other properties are associated with ***H_2O***. We can also put this point in terms of the epistemic intensions of *sentences* (functions from epistemic possibilities to truth values): ***Water is water*** is cognitively *in*significant because its epistemic intension is the constant function that takes every epistemic possibility to the True; ***Water is H_2O*** is cognitively significant because its epistemic intension takes certain epistemic possibilities to the True and others to the False.

Lastly, since these properties fill the reference-fixing role for ***water***, ***water*** refers to the unique clear, drinkable stuff found in the oceans and lakes. If some epistemic possibility says that XYZ is the unique stuff with these properties, then the epistemic intension of ***water*** will map that epistemic possibility to XYZ.

As is apparent from the above quotations, a competent speaker is supposed to be able to identify the referent of a word like ***water*** in an epistemic possibility E that is specified *without using the word* ***water*** (or cognate expressions): for example, a possibility in which "the clear, drinkable liquid in the oceans and lakes is XYZ." So competent speakers are not supposed simply to know that ***water*** refers to *water*. Likewise, a competent speaker is supposed to be able to identify the referent of ***Bob Dylan*** in epistemic possibilities that are specified without using the name ***Bob Dylan***. Let us put this point by saying that speakers are supposed to have *substantial identifying knowledge* of the referents of ***water*** and ***Bob Dylan***. This amounts to having an ability to evaluate, upon ideal a priori reflection, all instances of the schematic conditional (Turns-Out), where E is specified without using the word T (or any of its cognates).[8]

In fact, Chalmers thinks that speakers will be able to identify the referents of their words in epistemic possibilities specified in *strongly* reductive terms. The only expressive resources required, he thinks, are the language of a complete fundamental physics and a language suitable for describing "the phenomenal states and properties instantiated by every subject bearing such states and properties, at every time" (Chalmers and Jackson (2001), 319), plus a few other bells and whistles.[9] Given an epistemic possibility E specified using only these vocabularies, speakers who understand ***Bob Dylan*** and ***water*** are supposed to be able to identify the referents of ***Bob Dylan*** and ***water*** in E.

For our purposes, though, two-dimensionalism need not be viewed as having such strong reductive aspirations. We will just take the two-dimensionalist to be employing *some kind* of "reductive" specification of epistemic possibilities, leaving the details open.

So far we have given the impression that a word has a *unique* epistemic intension. However, the two-dimensionalist can and typically will allow that a word's epistemic intension often varies from speaker to speaker. For example, Chalmers considers the case of two speakers, who "have been exposed to different forms of water: one has only been exposed to water in liquid form (knowing nothing of a solid form), and the other has been exposed only to water in solid form (knowing nothing of a liquid form)" (2002b, 174). It might be, he says, that the epistemic intension of ***water*** as used by the first speaker differs from the epistemic intension of ***water*** as

[8] Three points of clarification. First, substantial identifying knowledge is intended to be *nothing more* than the ability to evaluate these conditionals. Chalmers and Jackson stress that this ability need not always be underwritten by the subject's explicit judgements about what properties T's referent possesses; often, they think, the ability will precede and explain any such judgements (see Chalmers and Jackson (2001), §3, and Jackson (1998b), 211–12).

Second, at the beginning of this paper we said that a speaker "associates P with T" iff the speaker believes that the referent of T (if it exists) has P. We should emphasize (again) that these beliefs may be ones that the subject has only "implicitly," in virtue of having the ability to evaluate these conditionals.

Finally, for our purposes, nothing turns on exactly how the notion of "ideal a priori reflection" is to be understood.

[9] The additions are a " 'that's all' statement" (Chalmers and Jackson (2001), 317) and a " 'you are here' marker" (318).

used by the second, although of course both intensions return the same referent at the actual world, namely H_2O. Again, to accommodate Putnam's ***elm*/*beech*** example, Chalmers says that the epistemic intension of ***elm*** as used by the botanical ignoramus is (roughly) given by the description ***The tree the experts call 'elm'*** while the epistemic intension of ***elm*** as used by the experts is something quite different (2002a, 617–18). Since none of our arguments turns on the assumption that words have unique epistemic intensions, for convenience we will mostly ignore this kind of alleged variation.

It is a strong and unobvious claim that speakers have substantial identifying knowledge of the referents of words like ***water*** and ***Bob Dylan***. Why think that they do? If speakers *do not* have this identifying knowledge, then they will not be in a position to know what these words refer to in the two-dimensionalist's reductively specified epistemic possibilities, and hence the corresponding epistemic intensions will not be well-defined. So another way of asking our question is: why think that words like ***water*** and ***Bob Dylan*** have epistemic intensions of the sort we have described?

3. Chalmers' Argument from Examples

There are various arguments for two-dimensionalism in the literature. Some of these are of an indirect sort: two-dimensionalism should be accepted because it neatly solves some theoretical puzzles—for example, puzzles about the necessary a posteriori.

Other arguments are more direct. For example, Chalmers says that two-dimensionalism is suggested naturally by armchair reflection on what speakers would say if the world turned out one way rather than another. In this way speakers can manifest their alleged abilities to identify the referents of words in different epistemic possibilities.

In Section 2, we quoted a few passages from Chalmers that are intended to exhibit a fragment of the epistemic intensions of ***Hesperus*** and ***water***. In the second of those passages, Chalmers suggests that the following conditional is a priori (that is, it can be known to be true by anyone who understands it, after a priori reflection):

> (CDL) If it turns out that XYZ is the clear, drinkable liquid in the oceans and lakes, then it will turn out that water is XYZ.

The *Argument from Examples*, as we will call it, starts with a discussion of ***water*** and other examples, and concludes that "[t]he intuitive characterization of epistemic intensions using the heuristics I have given here makes a strong prima facie case that expressions have epistemic intensions" (2002b, 146).

Now if (CDL) really is a priori, then this would help support a crucial part of the two-dimensional package, namely that speakers have what we called *substantial identifying knowledge* of the referents of their words. And perhaps with further argument, it can be used to support all the main two-dimensional claims. So, is (CDL) a priori?

Offhand, it can appear that way. Admittedly, given the present state of chemical knowledge, it would be somewhat deviant to utter (CDL) assertively. But we can

imagine some chemical ignoramus justifiably doing so, and it seems that the sentence she utters is *true*. (After all, it *has not* turned out that XYZ is the clear, drinkable liquid, etc. When the ignoramus discovers that water is H_2O, she does not have to *retract* her earlier assertion of (CDL).) Presumably the ignoramus could even *know* that (CDL) is true. And since she is ignorant, it might seem that in order to know that (CDL) is true, she only needs to understand it.

But consider an obvious fact about water, for instance that it is the liquid that comes out of taps in Barcelona, and consider the conditional:

(TAPS) If it turns out that XYZ is the liquid that comes out of taps in Barcelona, then it will turn out that water is XYZ.

Just as before, it would be somewhat deviant to utter (TAPS) assertively, but a chemical ignoramus might well do so. Again as before, it seems that (TAPS) is true, and that the ignoramus could know this to be so.

However—we may safely presume—(TAPS) is not a priori. When we imagine the ignoramus assertively uttering (TAPS), we are tacitly assuming that she knows some obvious a posteriori facts about water, in particular that it comes out of taps in Barcelona. If we imagine instead that the ignoramus has never heard of Barcelona, or that she believes that wine comes out of Barcelona taps, then she will have no justification for uttering (TAPS).

This should raise considerable suspicion concerning the status of (CDL). The fact that it is easy to imagine a scientific ignoramus knowing (CDL) to be true does not support the claim that the conditional is a priori. For in imagining the ignoramus to know (CDL), we may be tacitly assuming that she knows some obvious a posteriori facts about water, in particular that it is the clear, drinkable liquid in the oceans and lakes.

Now it might be insisted that even if we explicitly *stipulate* that the ignoramus has no a posteriori knowledge (beyond that conferred by her knowledge of English), it will *still* be plausible that she would be justified in accepting (CDL). Well, perhaps. Our only point at present is that one tempting but superficial reason for thinking that (CDL) is a priori collapses on further examination.

Having made this defensive point, it is time to go on the offensive. We think that familiar arguments from Kripke and Putnam show quite conclusively that no conditional like (CDL) is a priori. More-or-less equivalently, they show that speakers do not ordinarily have substantial identifying knowledge of the referents of words. Let us turn, then, to these arguments; in particular, to Kripke's arguments from ignorance and error.

4. Kripke's Arguments from Ignorance and Error

Kripke's examples of the names ***Cicero*** and ***Feynman*** support the view that a speaker can be a competent user of a name despite lacking substantial identifying knowledge *because of ignorance*. "[M]ost people," Kripke says, "when they think of Cicero, just think of *a famous Roman orator*, without any pretension to think that either there was

only one famous Roman orator or that one must know something else about Cicero to have a referent for the name" (1980, 81). Similarly, the man in the street may use the name ***Feynman*** to refer to Feynman, even though "[w]hen asked he will say: well, he's a physicist or something. He may not think this picks out anyone uniquely." (1980, 81)

Kripke's story about Gödel and Schmidt supports the view that a speaker can be a competent user of a name despite lacking substantial identifying knowledge *because of error*. In Kripke's story, speakers use the name ***Gödel*** to refer to Gödel, even though the achievements they ascribe to Gödel—discovering the incompleteness of arithmetic—were really performed by the unfortunate Schmidt. The properties that speakers associate with the name ***Gödel*** are rich enough to uniquely identify someone, but the person they uniquely identify is not the name's referent.

Notice that the Gödel/Schmidt story does not *just* teach us something about speakers who have false beliefs. It teaches us something stronger, namely that for *any* speaker (not just speakers in error), the properties the speaker associates with the name ***Gödel*** do not fill the reference-fixing role (with possible exceptions for those who named Gödel in the first place, or for the property *being Gödel*). For consider some competent user of the name ***Gödel*** who *knows* that it refers to the individual having such-and-such properties—say, the property of discovering the incompleteness of arithmetic. Since the speaker knows that the referent of ***Gödel*** has this property, she believes it does, and hence she *associates* this property with the name. However, if this property filled the reference-fixing role, then in a nearby possible world in which the Schmidt story is true, and the speaker uses the name ***Gödel*** with the same semantic intentions, we should find that ***Gödel*** refers in her mouth to Schmidt. But for typical speakers, this is just what we do not find. Typical speakers may know that Gödel discovered the incompleteness of arithmetic, but they do not give that property reference-fixing authority.

Similarly with ***water***. Most competent users of this word *do* know that it refers to the kind that has certain properties, for instance the kind many instances of which are clear, drinkable, liquid, and found in the oceans and lakes. But considerations just like those in the Gödel/Schmidt case show that these associated properties do not fill the reference-fixing role for ***water*** (as the word is used by these speakers).

The arguments from ignorance and error are concerned with a typical user of a name who has picked it up from someone else. It might be argued that associated properties will at least be needed to fill the reference-fixing role in the special case where a speaker explicitly introduces a name. This is an issue too large to be properly discussed here, but it is worth noting that the matter is not at all straightforward. Take the case of ostensive definition. Suppose a speaker sees a dog, and dubs him ***Checkers***. There will be many properties that pick out the dog (for example, *being the dog the speaker is looking at*). But it is unclear whether the speaker needs to associate any such properties with the word, and a fortiori unclear whether any such properties fill the reference-fixing role. And even if an associated property *does* fill the reference-fixing role, it might be the unexciting property of *being this particular dog, Checkers*. The speaker may be able to name the dog ***Checkers*** simply because she stipulates, *of* the dog she is seeing, that *it* is the referent of ***Checkers***. These sorts of associated properties

seem ill-suited for Chalmers' purposes; they will not provide the kind of substantial identifying knowledge that he is looking for.

In any case, concentrating on typical speakers, the arguments from ignorance and error seem to show that associated properties do not fill the reference-fixing role for words like ***water*** and ***Bob Dylan***. Therefore these words, as used by typical speakers, do not have epistemic intensions. Chalmers, though, is quite unimpressed by these arguments, for reasons that we will now examine.

5. Chalmers' Response to Kripke

The core of Chalmers' response to the arguments from ignorance and error is expressed in the following passage:

> Does this argument against the description theory [that is, Kripke's arguments from ignorance and error] yield an argument against the intensional framework I have been outlining? It seems clear that it does not. This argument works with a conception of descriptions on which they correspond to linguistic expressions. When Kripke argues that the descriptions that the speaker "associates with" the name cannot fix reference, he always invokes linguistic descriptions that the speaker associates with the name, or at least explicit descriptive beliefs of the speaker. But the intensional framework is not committed to the idea that descriptions always correspond to linguistic expressions; in fact, at least part of the motivation of the framework comes from an independent rejection of this idea. And the intensional framework is not even committed to the idea that the intensions associated with a name correspond to explicit beliefs of the speaker. So there is no clear argument against the intensional framework here.
>
> In fact, Kripke's central method of argument seems to be obviously compatible with the intensional framework. A proponent of this framework could cast the argument strategy as follows. We want to show that for a given name N and description D, 'N is D' is not a priori. To do this, we consider a specific epistemically possible scenario W. We then reflect on a question such as the following: 'if W turns out to be actual, will it turn out that N is D?' And we find that the answer is no. If so, the epistemic intension of 'N is D' is false in W. So 'N is D' is not a priori.
>
> On this interpretation, when we think about the Gödel/Schmidt case, for example, we are tacitly evaluating the epistemic intension of 'Gödel' at a world as specified in the example. When we consider that world as an epistemic possibility, it reveals itself as an instance of the epistemic possibility that Gödel did not discover incompleteness. That is, we find that the epistemic intension of 'Gödel' does not pick out the prover in this world, it picks out the publisher. If so, the epistemic intensions of 'Gödel' and of 'the man who discovered the incompleteness of arithmetic' are distinct. (2002b, 169)

There are three main points in this passage. First, as Chalmers puts it a little later, "Kripke's arguments suggest that the epistemic intension of a name such as 'Gödel' cannot be precisely captured in a linguistic description. But they do nothing to suggest that the epistemic intension does not exist" (2002b, 170). Second, even if the description is linguistically expressible, the speaker might associate it with the name only tacitly or implicitly—if asked for an explicit statement of what properties she was using to identify the referent of ***Gödel*** in various epistemic possibilities, she might be at a loss. Third, Kripke's own methodology is best viewed as a way of *revealing* or *articulating* a name's epistemic intension, rather than as demonstrating that the

name *has no* epistemic intension. (See Chalmers (2002a), n. 11; Chalmers and Jackson (2001), 326–7; and Jackson (1998b), 212–14.)

Take the first point first. Suppose that a speaker has seen a proof of the first incompleteness theorem, and retains a capacity to recognize the proof visually. Let us further suppose that the speaker associates some properties with the name ***Gödel*** that she cannot fully articulate in English. The best she can do is something like ***the man who discovered* this *proof***, uttered while demonstrating the appropriate pages in *From Frege to Gödel: A Source Book in Mathematical Logic*. But *what* she has in mind is an essentially visual way of thinking of the proof; her demonstrative utterance (let us suppose) does not fully articulate it.

Kripke's story about Schmidt straightforwardly shows that these associated properties do not fill the reference-fixing role for ***Gödel***, as it is used by this speaker. For, in the story, the person who possesses these properties is Schmidt; yet the speaker's word ***Gödel*** refers to Gödel. Further, any other linguistically inexpressible properties that a speaker might associate with ***Gödel*** would also appear to be subject to a Schmidt-type objection. So although Chalmers is right to claim that the properties that a speaker associates with ***Gödel*** need not be linguistically expressible, this does not seem to help at all in fending off Kripke's argument from error.

Neither does the first point help in fending off Kripke's argument from ignorance. Perhaps many ordinary speakers have some complex idea of ancient Rome, derived from *Ben Hur* and *Gladiator*, that resists complete articulation in English. The properties they associate with ***Cicero*** might be gestured at with phrases like ***a famous orator from* that place**, while demonstrating various sword-and-sandal scenes. So the properties they associate with ***Cicero***, let us suppose, are also not linguistically expressible. But obviously these properties do not pick out the referent of ***Cicero*** uniquely. And since there is no reason to suppose that ordinary speakers associate other linguistically inexpressible properties with ***Cicero*** that *do* pick out the referent uniquely, the argument from ignorance stands.

Turn now to Chalmers' second point, the one about explicitness. This point does indicate a need for caution: we should not conclude that an ordinary speaker does not have substantial identifying knowledge of the referent of ***Cicero*** just because the speaker herself cannot explicitly state it. Substantial identifying knowledge might make its presence known through the speaker's disposition to apply the name, rather than through her verbal reports (see note 8). But it seems clear that even when we take this into account, ordinary speakers are often impressively ignorant about the referents of names like ***Cicero***. Their poor performance on history exams is due to their *lack* of knowledge of Cicero's life and times, not to its *implicitness*. And in any case, even if we found that speakers did associate properties with ***Cicero*** that were both suitably reductive and uniquely identifying, the Gödel/Schmidt example shows that they usually will not fill the reference-fixing role.

Finally, let us turn to Chalmers' third point, that Kripke's examples help to *reveal* or *articulate* a name's epistemic intension, rather than demonstrate that it does not have one. Recall that the epistemic intension of ***Gödel*** is supposed to represent a speaker's ability to identify the referent of ***Gödel*** in some *reductively specified* epistemic possibility—her *substantial identifying knowledge*. Possibilities specified as

ones containing *Gödel* do not count. So Chalmers seems to be saying that evaluating Kripke's example involves identifying the referent of ***Gödel*** in a reductively specified epistemic possibility. Kripke gives us an epistemic possibility in which certain people (bearing the names ***Schmidt*** and ***Gödel***) do certain things, and given that epistemic possibility, "the epistemic intension of 'Gödel' does not pick out the prover [of the theorem] in this world, it picks out the publisher." If that is the right account of Kripke's Gödel/Schmidt example, then it would *not* show that ***Gödel*** lacks an epistemic intension. Rather, the example would presuppose that ***Gödel*** *has* an epistemic intension, and it would help us to articulate what that intension is.

However, we think this is a misrepresentation of Kripke's example. *Kripke* does not offer any reductive specification of the Gödel/Schmidt possibility. As we read Kripke, he is asking us, in effect, to imagine a situation in which a speaker who falsely believes that ***Gödel*** refers to the man who discovered the incompleteness of arithmetic, nonetheless *uses* ***Gödel*** *to refer to Gödel*. The situation is specified in terms of properties that *Gödel* does and does not have. ("A man named 'Schmidt' . . . actually did the work in question. His friend Gödel somehow got hold of the manuscript . . ." (Kripke 1980, 84).) Kripke's point is that that situation is perfectly coherent, which makes it plausible that the referent of the name ***Gödel*** is not fixed by properties that the speaker associates with it. There is nothing at all in Kripke's description of the example to support the view that a competent user of the name ***Gödel*** can identify its referent in some *reductively specified* epistemic possibility. So there are no grounds here for thinking that anyone who understands ***Gödel*** has substantial identifying knowledge about its referent.[10]

Kripke *could have* presented his story about Gödel and Schmidt without using the name ***Gödel***, but by using instead an expression his readers knew to apply to Gödel, such as ***the member of the Institute for Advanced Study who starved himself to death***. If he had done so, though, he would have been exploiting shared a posteriori identifying knowledge about Gödel, rather than identifying knowledge that we all have just in virtue of understanding ***Gödel***.

It may be that anyone who understands ***Gödel*** will know some substantial conditions that are *necessary* for being the referent of ***Gödel***. (Conditions, that is, that can be specified without using ***Gödel*** or its cognates.) For example, perhaps anyone who understands ***Gödel*** knows that it refers to a sentient being, if it refers at all. If so, the conditional ***If it turns out that there are no sentient beings, then it will turn out that Gödel does not exist*** will be a priori. Competent speakers may also know some interesting *sufficient* conditions for being the referent of ***Gödel***. For example, if competent speakers know the necessary condition just mentioned, then they will also know that if there is exactly one sentient being and if ***Gödel*** refers, then it refers to this sentient being. If so, the conditional ***If it turns out that Gödel exists and there is exactly one sentient being, then it will turn out that Gödel is this sentient being*** will be a priori.

It is, however, a considerably stronger claim that competent speakers know substantial conditions that are *both necessary and sufficient* for being the referents of

[10] Soames (2005) offers similar objections.

their terms; that is, substantial identifying knowledge. We do not think that examples like Kripke's provide any support for this strong claim—even if the knowledge is allowed to be linguistically inexpressible and implicit.[11] Our ability to identify referents in such examples typically owes to the fact that the examples are specified in *non*-substantial terms, or are specified using descriptions that the referents are *known a posteriori* to satisfy, or both. So these examples do not give us reason to attribute substantial identifying knowledge. Rather, as Kripke says, they show that competent speakers do *not* typically need to have such knowledge.

It seems to us, then, that Chalmers' three points do not deflect the force of Kripke's ***Cicero*** and ***Gödel*** examples.

6. The Metalinguistic Response to Kripke

After responding to Kripke's arguments, Chalmers turns to the question of whether the epistemic intension of a name like ***Gödel*** can "at least be approximated by a linguistic description." "This is not compulsory for the intensional framework," he says, "but it can at least be enlightening to look" (2002b, 170).

> To answer this question, one needs to consider: when speakers use a name such as 'Gödel' or 'Feynman' in cases such as those above [that is, when they are mistaken or ignorant], how do they determine the referent of the name, given sufficient information about the world? For example, if someone knows only that Feynman is a famous physicist and that Gell-Mann is a famous physicist, how will external information allow her to identify the distinct referents of 'Feynman' and 'Gell-Mann'? The answer seems clear: she will look to *others'* use of the name. Further information will allow her to determine that members of their community use 'Feynman' to refer to a certain individual, and that they use 'Gell-Mann' to refer to a different individual. Once she has this information, she will have no problem determining that her own use of 'Feynman' refers to the first, and that her own use of 'Gell-Mann' refers to the second.
>
> This suggests that if we want to approximate the epistemic intension of the speaker's use of 'Feynman' in a description, one might start with something like 'the person called "Feynman" by those from whom I acquired the name.' It certainly seems that if relevant information about others' uses is specified in an epistemic possibility, then this sort of description will usually give the right results. The same goes for the 'Gödel' epistemic possibility. In all these cases, it seems that a name is being used *deferentially*: in using a name, the speaker defers to others who use the name. (2002b, 170–1, note omitted).

We think that this should be Chalmers' official response to the epistemological arguments, not the three points discussed above. The moral of the arguments from ignorance and error is that if two-dimensional account of names is to be workable, then the epistemic intension of a name like ***Gödel*** cannot be given by any sort

[11] Two-dimensionalists sometimes shy from claiming we know *necessary and* sufficient substantial conditions for being the referent of a name. They sometimes only talk about knowing substantial sufficient conditions. If our identifying knowledge (or the dispositions that constitute it—see note 8) is to play a reference-fixing role, then necessary and sufficient conditions seem to be needed. In any case, though, we take the doubts we're raising in this paper to apply to our having implicit a priori knowledge even of substantial sufficient conditions, except for special conditions of the sort we mention in the text.

of "famous deeds" description, like ***the man who discovered the incompleteness of arithmetic***. Instead, the epistemic intension has to be given by something like the description ***the person called 'Gödel' by those from whom I acquired that name***.[12]

As Chalmers notes, Kripke discusses various proposals along these lines, for example "By 'Gödel' I mean the man *Jones* calls 'Gödel' " (1980, 92). These proposals are said either to fall to a Gödel/Schmidt type objection, or else to violate Kripke's noncircularity requirement.

And, again as Chalmers notes, his own proposal seems to be vulnerable to Gödel/Schmidt-type objections. To accommodate cases where the speaker mishears or misremembers the name, Chalmers tries a "closer approximation":

> Perhaps 'The referent of the relevant name used by the person from whom I acquired the antecedent of my current term "Gödel" ' would do a better job. But no doubt there would be further counterexamples . . . But as in all these cases, the most this shows is that any such approximation is imperfect. One refutes these approximations by evaluating the epistemic intension in certain epistemic possibilities and showing that the approximation give the wrong results; so this sort of argument does nothing to show that the epistemic intension does not exist. (2002b, 171)

Indeed, there are further counterexamples. Suppose a speaker baptizes Gödel with the name ***Gödel***, and so does not acquire the name from someone else. Further, suppose she forgets that this is so. Her use of ***Gödel*** still refers to Gödel. But if the property *being the referent of the relevant name used by the person from whom she acquired the antecedent of her current term* ***Gödel*** filled the reference-fixing role, then—since she never acquired ***Gödel*** from anyone—her use of ***Gödel*** would not refer.

Even if we assume that the metalinguistic proposal can be fixed up to avoid obvious counterexamples, at least three objections remain.

The first objection is that the metalinguistic proposal imposes unreasonable demands on understanding a word. Admittedly, the proposal does not require speakers to have explicit metalinguistic beliefs (see note 8 and the preceding section). But it does require competent speakers to have *an ability to evaluate conditionals* whose antecedents contain sophisticated semantic vocabulary, like ***the antecedent of my current term n, the referent of a term as used by speaker S***, and so on. One would have thought, on the contrary, that the ability to speak and understand a language comes *first*: understanding words is a precondition of such conceptually sophisticated abilities, not the other way around.[13]

The second objection is that metalinguistic properties, even if they do fill the reference-fixing role, will not generally fill the Frege role. Consider an example. Imagine that Rosa Zola was taken to the high school prom by Robert Zimmerman; despite having a wonderful evening, they lost touch after graduation. One day many years later Rosa hears an assertive utterance of ***Bob Dylan is Robert Zimmerman***. She

[12] Note that Chalmers is allowing *semantic* specifications of epistemic possibilities here: for example, descriptions of the referential history of ***Gödel*** as used by a certain speaker. On his official reductive account, these are dispensable. See also Jackson (1998b), 209 ff. There is a large literature discussing metalinguistic proposals of this sort. Nelson (2002) gives a useful overview.

[13] For further discussion, see Braun (1995) and Soames (2005).

is utterly astonished and delighted. The information she gains is highly non-trivial, and it leads her to contrive a reunion with her old prom date. Two-dimensionalism promises an account of this: the cognitively significant information Rosa gains is the contingent proposition that the D is the Z, where *being D* determines the epistemic intension of ***Bob Dylan***, and *being Z* determines the epistemic intension of ***Robert Zimmerman***. However, on the metalinguistic proposal this contingent proposition is something of the following sort:

> The referent of ***Bob Dylan*** as used by those from whom Rosa acquired that name is the referent of ***Robert Zimmerman*** as used by those from whom Rosa acquired that name.

And this information is patently *not* the news that excited Rosa and moved her to action. What excited her, we may suppose, is the information that the singer of *Mr. Tambourine Man* is the person she dated in high school. Rosa gained this information by hearing ***Bob Dylan is Robert Zimmerman*** because she associates ***Bob Dylan*** with the property *being the singer of* Mr. Tambourine Man. The associated properties that play the Frege role will be "famous deeds" properties like this one, not metalinguistic properties. And as we have already argued, these "famous deeds" properties will typically be ill-suited to play the reference-fixing role. For typical speakers, those kinds of properties will always be vulnerable to Gödel/Schmidt-type counterexamples.[14]

So, adopting the metalinguistic proposal prevents epistemic intensions from solving Frege's problem, and thus removes one of the advertised advantages of two-dimensionalism. (See (Chalmers 2002a), 622–4; cf. Jackson (1998a), 76.)

The third objection is both the simplest and the most fundamental: the metalinguistic proposal is unmotivated. Before trying to make the metalinguistic proposal work, better reason is needed for thinking that the referent of a word *must* always be determined by the speaker's giving reference-fixing authority to some associated properties. In our opinion, no adequate case for this assumption has yet been supplied.

References

Braun, D. (1995). Katz on names without bearers. *Philosophical Review* 104: 553–76.

Byrne, A. (1999). Cosmic hermeneutics. *Philosophical Perspectives* 13: 347–83. <http://web.mit.edu/abyrne/www/cosmichermeneutics.pdf>

Chalmers, D. (2002a). The components of content (revised version). In *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers. Oxford University Press. <http://consc.net/papers/content.html>

____ (2002b). On sense and intension. *Philosophical Perspectives* 16: 135–82. <http://consc.net/papers/intension.html>

[14] This is the "other way," alluded to in Section 1 above, of making the point that properties that fill the Frege role need not fill the reference-fixing role. Here we are indebted to Thau (2002, ch. 3).

——— (2004). The foundations of two-dimensional semantics. This volume, Chapter 4.
——— and F. Jackson (2001). Conceptual analysis and reductive explanation. *Philosophical Review* 110: 315–60. <http://consc.net/papers/analysis.html>
Jackson, F. (1998a). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
Jackson, F. (1998b). Reference and description revisited. *Philosophical Perspectives* 12: 201–18.
Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.
Nelson, M. (2002). Descriptivism defended. *Noûs* 36: 408–35.
Pryor, J. (2003). Varieties of two-dimensionalism. <http://jimpryor.net/research/notes/Varieties2D.pdf>
Soames, S. (2005). Reference and descriptions. In *The Oxford Handbook of Contemporary Philosophy*, ed. F. Jackson and M. Smith. Oxford: Oxford University Press.
Stalnaker, R. (2001). On considering a possible world as actual. *Aristotelian Society Supplemental Volume* 75: 141–56.
Thau, M. (2002). *Consciousness and Cognition*. Oxford: Oxford University Press.

4

The Foundations of Two-Dimensional Semantics

David J. Chalmers

1. Meaning, Reason, and Modality

Why is two-dimensional semantics important? One can think of it as the most recent act in a drama involving three of the central concepts of philosophy: meaning, reason, and modality. First, Kant linked reason and modality, by suggesting that what is necessary is knowable a priori, and vice versa. Second, Frege linked reason and meaning, by proposing an aspect of meaning (sense) that is constitutively tied to cognitive significance. Third, Carnap linked meaning and modality, by proposing an aspect of meaning (intension) that is constitutively tied to possibility and necessity.

Carnap's proposal was intended as something of a vindication of Frege's. Frege's notion of sense is somewhat obscure, but Carnap's notion of intension is more clearly defined. And given the Kantian connection between reason and modality, it follows that intensions have many of the properties of Fregean senses. In effect, Carnap's link between meaning and modality, in conjunction with Kant's link between modality and reason, could be seen as building a Fregean link between meaning and reason. The result was a golden triangle of constitutive connections between meaning, reason, and modality.

Some years later, Kripke severed the Kantian link between apriority and necessity, thus severing the link between reason and modality. Carnap's link between meaning and modality was left intact, but it no longer grounded a Fregean link between meaning and reason. In this way the golden triangle was broken: meaning and modality were dissociated from reason.

Two-dimensional semantics promises to restore the golden triangle. While acknowledging the aspects of meaning and modality that derive from Kripke, it promises to explicate further aspects of meaning and modality that are more closely tied to the

An abridged version of this paper appeared under the title 'Epistemic Two-Dimensional Semantics' in a special issue of *Philosophical Studies* in 2004. Portions of this paper have been presented at the conferences on Two-Dimensionalism in Barcelona and ANU, at the Pacific Division meeting of the APA, and at UC Berkeley and the University of North Carolina. I am grateful to the audiences on those occasions for feedback, and to Manuel Garcia-Carpintero and Daniel Stoljar for detailed comments on the paper.

rational domain. In particular it promises to look at the space of possibilities in a different way, and to erect a notion of meaning on that basis. In this way, we might once again have a grip on an aspect of meaning that is constitutively tied to reason.

To date, this restoration has been incomplete. Many different ways of understanding two-dimensional semantics have been proposed, and many of them restore the triangle at best partially. It is controversial whether two-dimensional semantics can be understood in such a way that the triangle is fully restored. To see this is possible, we need to investigate the foundations of two-dimensional semantics, and explore the many different ways in which the framework can be understood. I think that when the framework is understood in the right way, it can reinstantiate the links between meaning, reason, and modality.

1.1 Frege, Carnap, and Kripke

We can begin with some more detailed background. If we squint at history from just the right angle, focusing on one strand of thought and setting aside others, we obtain a simplified rational reconstruction that brings out the key points.

It is useful to start with Frege. Frege held that an expression in a language typically has a *referent*—or what I will here call an *extension*. The extension of a singular term is an individual: for example, the extension of the name 'Hesperus' is the planet Venus, and the extension of the description 'the teacher of Aristotle' is Plato. The extension of a general term is a class. And the extension of a sentence is its truth-value.

Frege noted that the extension of an expression does not in general determine its cognitive significance: the role it plays in reasoning and in knowledge. For example, 'Hesperus' (the name used for the evening star) and 'Phosphorus' (the name used for the morning star) have the same referent but have different cognitive significance, as witnessed by the fact that 'Hesperus is Hesperus' is cognitively trivial, while 'Hesperus is Phosphorus' is nontrivial. The same goes for many other pairs of expressions: perhaps 'renate' (creature with a kidney) and 'cordate' (creature with a heart), or 'water' and 'H_2O', or 'I' (as used by me) and 'David Chalmers'. In each pair, the members are co-extensive (they have the same extension), but they are cognitively and rationally distinct.

Frege held that meaning is tied constitutively to cognitive significance, so that if two expressions have different cognitive significance, they have different meaning. It follows that there must be more to meaning than extension. Frege postulated a second aspect to meaning: *sense*. When two expressions are cognitively distinct, they have different senses. For example, the nontriviality of 'Hesperus is Phosphorus' entails that although 'Hesperus' and 'Phosphorus' have the same extension, they have a different sense. We can put the general idea as follows:

> **Fregean Thesis**: Two expressions 'A' and 'B' have the same sense iff 'A $\equiv$ B' is cognitively insignificant.

Here, 'A $\equiv$ B' is a claim that is true if and only if 'A' and 'B' have the same extension. Where 'A' and 'B' are singular terms, this will be the identity 'A = B'; where 'A' and 'B' are sentences, this will be the material biconditional 'A iff B'; and so on. As for cognitive significance, we can say at a first approximation that a claim is cognitively

insignificant when it can be known trivially by a rational being. As such, we can see this characterization of sense as providing a first bridge between meaning and reason.

The idea that expressions have senses is attractive, but senses are nevertheless elusive. What exactly is a sense? What exactly is cognitive significance? How does one analyze meanings beyond extensions? In the middle part of the twentieth century, a number of philosophers, notably Carnap, had an insight. We can use the notions of *possibility* and *necessity* to help understand meaning, and in particular to help understand sense.

There are many possible ways the world might be; and we can use language to describe these possibilities. An expression can be applied to the actual state of the world, yielding an actual extension, or it can be applied to alternative possible states of the world, yielding alternative possible extensions. Take expressions such as 'renate' and 'cordate'. In the world as it actually is, all renates are cordates, so these terms have the same extension. But it is not *necessary* that all renates are cordates: if the world had been different, some renates might have failed to be cordates. Applied to such an alternative possibility, the two terms have a different extension. We can say: 'renate' and 'cordate' are co-extensive, but they are not *necessarily* co-extensive. Carnap suggested that we say two expressions have the same *intension* if and only if they are necessarily co-extensive.[1] So 'renate' and 'cordate' have the same extension, but different intensions. We can put the general claim as follows:

Carnapian Thesis: 'A' and 'B' have the same intension iff 'A $\equiv$ B' is necessary.

What exactly is an intension? Carnap's characterization suggests a natural definition: an intension is a function from possibilities to extensions.[2] The possibilities here correspond to different possible states of the world. Relative to any possibility, an expression has an extension: for example, a sentence (e.g. 'All renates are cordates') can be true or false relative to a possibility, and a singular term (e.g. 'the teacher of Aristotle') picks out an individual relative to a possibility. An expression's intension is the function that maps a possibility to the expression's extension relative to that possibility. When two expressions are necessarily co-extensive, they will pick out the same extension relative to all possibilities, so they will have the same intension. When two expressions are not necessarily co-extensive, they will not pick out the same extension relative to all possibilities, so they will have different intensions. So intensions behave just as Carnap suggests they should.

[1] See Carnap (1947). The idea is also present in Lewis (1944).

[2] This definition of an intension is often attributed to Carnap, but in Carnap (1947) it plays at most a minor role. He proposes something like this (in section 40, p. 181) as a way of understanding individual concepts, which are the intensions of names, but then moves to a slightly different understanding. Earlier in the book, he characterizes necessity ("L-truth") in terms of state-descriptions, which are akin to possible worlds. But state-descriptions soon drop out of the discussion, so that intensions are treated in effect as something of a primitive semantic value. This sort of construction is also discussed in Carnap (1963), 892–94. (Thanks to Wolfgang Schwartz for pointers here.) A proposal close to the definition above is present in C. I. Lewis's suggestion that an intension "comprises whatever must be true of any possible world in order that the proposition should apply to it or be true of it" (Lewis 1944).

Seen this way, the notion of an intension provides a bridge between meaning and modality. Just as a sense can be seen as a sort of meaning that is constitutively tied to reason, an intension can be seen as a sort of meaning that is constitutively tied to modality. Furthermore, intensions seem to behave very much as senses are supposed to behave. Just as two expressions can have the same extension but different senses, two expressions can have the same extension but different intensions. And just as sense was supposed to determine extension, intension seems to determine extension, at least relative to a world.

One can make a direct connection by adding an additional claim connecting modality and reason. It has often been held that a proposition is necessary if and only if it is a priori (knowable independently of experience) or trivial (yields no substantive knowledge of the world). The notions of apriority and triviality are essentially rational notions, defined in epistemic terms. Carnap himself held a version of the thesis involving triviality, but it is more useful for our purposes to focus on the version involving apriority. In this form, the relevant thesis goes back at least to Kant, so we can call it:

> **Kantian Thesis**: A sentence S is necessary iff S is a priori.

If we combine the Carnapian Thesis with the Kantian Thesis, we obtain the following:

> **Neo-Fregean Thesis**: Two expressions 'A' and 'B' have the same intension iff 'A $\equiv$ B' is a priori.

If this claim is accepted, then one has recaptured something that is at least close to the Fregean thesis. For apriority is at least closely related to cognitive insignificance. When a proposition is cognitively insignificant, it is plausibly a priori. The reverse is not the case, on Frege's understanding of cognitive significance: many logical and mathematical propositions are cognitively significant, even though they are a priori. But in any case, apriority and cognitive insignificance are at least closely related rational notions. Typical cognitively significant identities, such as 'Hesperus is Phosphorus', 'Water is H_2O', and 'I am David Chalmers' are all a posteriori. If the Neo-Fregean Thesis is correct, it follows that 'Hesperus' and 'Phosphorus' have different intensions, as do 'water' and 'H_2O', and 'I' and 'David Chalmers'. So intensions behave quite like Fregean senses.

In effect, modality serves as a bridge in explicating the tie between meaning and reason. One constructs a notion of meaning using modal notions, combines this with the claim that modality is constitutively tied to reason, and ends with a link between all three. The central connection between meaning, reason, and modality is captured within the Neo-Fregean thesis: intension is a notion of meaning, defined in terms of modality, that is constitutively connected to reason.

This golden triangle was shattered by Kripke, who cut the connection between reason and modality. Kripke argued that the Kantian Thesis is false: there are many sentences that are necessarily true but whose truth is not knowable a priori. For example, Kripke argued that given that Hesperus is actually Phosphorus, it could not have been that Hesperus was not Phosphorus: Hesperus is necessarily the planet

Venus, and so is Phosphorus. So although 'Hesperus is Phosphorus' is not knowable a priori, it is nevertheless necessary. More generally, Kripke argued that names and natural kind terms are rigid designators, picking out the same extension in all possible worlds. It follows that any true identity involving such terms is necessary. For example, 'Water is H_2O' is necessary, even though it is a posteriori. The same goes for claims involving indexicals: 'I am David Chalmers' (as used by me) is another a posteriori necessity.

If Kripke is right about the Kantian Thesis, then the Neo-Fregean Thesis is also false. Since 'Hesperus is Phosphorus' is necessary, 'Hesperus' and 'Phosphorus' have the same intension, picking out the planet Venus in all possibilities. But the equivalence between 'Hesperus' and 'Phosphorus' is nevertheless a posteriori and cognitively significant. So cognitively and rationally distinct pairs of expressions can have the same intension: witness 'Hesperus' and 'Phosphorus', 'water' and 'H_2O', 'I' and 'David Chalmers'. So the Neo-Fregean Thesis fails, and intensions no longer behave like Fregean senses.

In effect, Kripke leaves intact the Carnapian link between meaning and modality, but in severing the Kantian link between reason and modality, he also severs the Fregean link between meaning and reason. This is roughly the received view in contemporary analytic philosophy: meaning and modality are connected, but both are disconnected from reason.

1.2 Two-dimensional semantics

Although most contemporary analytic philosophers accept Kripke's arguments against the Kantian thesis, many would still like to hold that Frege was right about *something*. There remains an intuition that 'Hesperus' and 'Phosphorus' (or 'water' and 'H_2O', or 'I' and 'David Chalmers') differ in at least some dimension of their meaning, corresponding to the difference in their cognitive and rational roles. One might try to do this by breaking the Carnapian connection between meaning and modality. Two-dimensional semantics takes another strategy: in effect, it finds another way of looking at modality that yields a Fregean aspect of meaning.

The core idea of two-dimensional semantics is that there are two different ways in which the extension of an expression depends on possible states of the world. First, the actual extension of an expression depends on the character of the actual world in which an expression is uttered. Second, the counterfactual extension of an expression depends on the character of the counterfactual world in which the expression is evaluated. Corresponding to these two sorts of dependence, expressions correspondingly have two sorts of intensions, associating possible states of the world with extensions in different ways. On the two-dimensional framework, these two intensions can be seen as capturing two dimensions of meaning.

These two intensions correspond to two different ways of thinking of possibilities. In the first case, one thinks of a possibility as representing a way the actual world might turn out to be: or as it is sometimes put, one *considers a possibility as actual*. In the second case, one acknowledges that the actual world is fixed, and thinks of a possibility as a way the world might have been but is not: or as it is sometimes put, one *considers a possibility as counterfactual*. When one evaluates an expression relative

to a possible world, one may get different results, depending on whether one considers the possible world as actual or as counterfactual.

The second way of thinking about possibilities is the more familiar in contemporary philosophy. Kripke's arguments rely on viewing possibilities in this way. Take a possibility in which the bright object in the evening sky is a satellite around the earth, and in which Venus is visible and bright only in the morning. When we think of this possibility as a counterfactual way things might have been, we do not describe it as a possibility in which Hesperus is Mars, but as one in which Hesperus (and Phosphorus) is invisible in the evening. So relative to this possibility considered as counterfactual, 'Hesperus' picks out Venus. Correspondingly, the second-dimensional intensions of both 'Hesperus' and 'Phosphorus' both pick out Venus in this possibility, and in all possibilities in which Venus exists. It is this familiar sort of intension that yields the Kripkean gap between intension and cognitive significance.

The first way of thinking about possibilities is the less familiar in contemporary philosophy. If we take the possibility described above, and think of it as a way the world might actually be, we can say: if the world really is that way, then 'Hesperus' picks out a satellite. So relative to this possibility considered as actual, 'Hesperus' picks out not Venus but the satellite. Correspondingly, the first-dimensional intension of 'Hesperus' picks out the satellite in this possibility, while that of 'Phosphorus' picks out Venus. So 'Hesperus' and 'Phosphorus' have different first-dimensional intensions. This difference is tied to the fact that the actual-world reference of 'Hesperus' and 'Phosphorus' is fixed in quite different ways, although as things turn out, their referents coincide. Because of this, it seems that the first dimension may be better suited than the second for a link to reason and to cognitive significance.

The possibilities evaluated in the second dimension are usually thought of as possible worlds. The possibilities evaluated in the first dimension are a little different, as they reflect the nature of a world from the point of view of a speaker using an expression within a world. It is useful for many purposes to see these possibilities as *centered worlds*: worlds marked with a "center", which is an ordered pair of an individual and a time. We can think of the center of the world as representing the perspective of the speaker within the world.

I have been deliberately vague about just how the relevant intensions are to be defined, since as we will see, there are many different ways to define them. Because of this, giving detailed examples is tricky, because different frameworks treat cases differently. Nevertheless, it is useful to go through some examples, giving an intuitive analysis of the results that two-dimensional semantics might be expected to give *if* it is to yield something like a Fregean sense in the first dimension. We will later see how this can be cashed out in detail. For now, I will use "1-intension" as a generic name for a first-dimensional intension, and "2-intension" as a generic name for a second-dimensional intension.

First, 'Hesperus is Phosphorus'. In a centered world considered as actual, this is true roughly when the morning star visible from the center of that world is the same as the evening star. In a world considered as counterfactual, it is true when Venus is Venus. 'Hesperus' functions roughly to pick out the evening star in the actual world, so the 1-intension of 'Hesperus' picks out the evening star in a given centered

world. Likewise, the 1-intension of 'Phosphorus' picks out the morning star in a centered world. Both of these terms behave rigidly in counterfactual evaluation, so their 2-intensions pick out their actual referents in all worlds. So the 2-intensions of both 'Hesperus' and 'Phosphorus' pick out Venus in all worlds.

Second, 'Water is H_2O'. In a centered world considered as actual, this is true roughly when the clear, drinkable liquid around the center of that world has a certain pattern of chemical structure. In a world considered as counterfactual, it is true when H_2O is H_2O. The reference of 'water' is fixed roughly by picking out the substance with certain superficial properties and a certain connection to the speaker in the actual world, so its 1-intension picks out roughly the substance with those properties connected to the center of a given world. Similarly, the 1-intension of 'H_2O' picks out the substance with the right sort of chemical structure in a centered world. As in the first case, both expressions behave rigidly in counterfactual evaluation, so their 2-intensions pick out H_2O in all worlds.

Third, 'I am a philosopher'. In a centered world considered as actual, this sentence is true when the being at the center of the world is a philosopher. In a world considered as counterfactual, this sentence (or at least my utterance of it) is true if David Chalmers is a philosopher in that world. The actual-world reference of 'I' is fixed by picking out the subject who utters the token; so the 1-intension of 'I' picks out the subject at the center of a given world. 'I' behaves as a rigid designator in counterfactual evaluation, so its 2-intension picks out the actual referent (in this case, David Chalmers) in all possible worlds. 'Philosopher', by contrast, is a broadly descriptive term: both its 1-intension and its 2-intension function to pick out beings with certain characteristic attributes. Certain patterns seem to emerge. The first two sentences are necessary (at least if Kripke is right), and both of them have a 2-intension that is true in all worlds. The third sentence is contingent, and its 2-intension is false in some worlds. So it seems that a sentence is necessary precisely when it has a necessary 2-intension. This corresponds directly to the Carnapian thesis: 2-intensions, in effect, are defined so that two expressions will have the same 2-intensions when they are necessarily equivalent.

On the other hand, all three of these sentences are a posteriori, and all of them appear to have a 1-intension that is false in some centered worlds. At the same time, a priori sentences such as (perhaps) 'All bachelors are unmarried males' or 'Hesperus (if it exists) has been visible in the evening' can plausibly be seen as having a 1-intension that is true in all centered worlds. So it is at least tempting to say that a sentence is a priori precisely when it has a necessary 1-intension. This corresponds to the neo-Fregean thesis: one might naturally suggest that two expressions have the same 1-intension precisely when they are a priori equivalent. To illustrate, one can note that the difference in the 1-intensions of 'Hesperus' and 'Phosphorus', or of 'water' and 'H_2O', seems to be closely tied to their a priori inequivalence. All this needs to be analyzed in more depth, but one might at least characterize the general sort of behavior suggested in the examples above, where differences in 1-intensions go along at least roughly with differences in cognitive significance, as *quasi-Fregean*.

Along with the 1-intension and the 2-intension of a given expression, one can also define a *two-dimensional intension*. In many cases, just as an expression's extension

depends on how the actual world turns out, an expression's 2-intension depends on how the actual world turns out. The expression's two-dimensional intension captures this dependence: it can be seen as a function from centered worlds to 2-intensions, or equivalently as a function from pairs of centered worlds and worlds to truth-values. In the case of 'Hesperus', for example, the two-dimensional intension maps a centered world V to the 2-intension that picks out V's evening star (if it exists) in any worlds W. The actual 2-intension of an expression corresponds to the two-dimensional intension evaluated at the actual centered world of the speaker: given that Venus is the actual world's evening star, the 2-intension of 'Hesperus' picks out Venus in all worlds. The 1-intension of an expression can be reconstructed by "diagonalizing" the two-dimensional intension: one evaluates the two-dimensional intension at a centered world W, yielding a 2-intension, and then one evaluates this 2-intension at the same world (stripped of its center). One might think of the two-dimensional intension as representing the way that an expression can be used to evaluate counterfactual worlds, depending on which world turns out to be actual.

1.3 Varieties of two-dimensional semantics

I will return to these themes later, but for now it must be acknowledged that the situation is much more complicated than I have made things sound. A number of different two-dimensional systems have been introduced, and many of these give different results. A partial list of proponents of these systems, along with the names they give to their two-dimensional notions, includes:

Kaplan (1978; 1989): character and content
Stalnaker (1978): diagonal proposition and proposition expressed
Evans (1979): deep necessity and superficial necessity
Davies and Humberstone (1981): "fixedly actually" truth and necessary truth
Chalmers (1996): primary intension and secondary intension
Jackson (1998a): A-intension and C-intension

There are many differences between these systems, some on the surface, and some quite deep. Surface differences include the fact that where Chalmers and Jackson speak of two sorts of intensions, Evans and Davies and Humberstone speak of two sorts of necessity, while Kaplan and Stalnaker speak of propositions. This sort of difference is mostly intertranslatable. Given a notion of necessity and a corresponding way of evaluating possibilities (as with Evans and Davies and Humberstone), one can define a corresponding sort of intension, and vice versa. Stalnaker's propositional content is just a set of possible worlds, which is equivalent to the intension of a sentence, and Kaplan's content is closely related.[3] Kaplan's

[3] Kaplan's content is strictly speaking a singular proposition rather than a set of worlds, but it immediately determines a set of worlds. For our purposes, the difference between singular propositions, other structured propositions, and sets of worlds in analyzing the second dimension of content will not be crucial, so for simplicity I will speak as if the relevant second-dimensional contents are intensions. The discussion can be straightforwardly adapted to other views.

and Stalnaker's first-dimensional notions are defined over contexts (which are at least closely related to centered worlds), and initially involve a two-dimensional intension: a function from contexts to 2-intensions. Stalnaker diagonalizes this function, yielding a function from contexts to truth-values, or a 1-intension. Kaplan leaves his character as a two-dimensional function from contexts to 2-intensions, but a corresponding step could straightforwardly be taken. So in all these cases, there is a similar formal structure.

At a conceptual level, the systems have something further in common. In each case, the first-dimensional notion is put forward at least in part as a way of better capturing the cognitive or rational significance of an expression than the second dimension. And in each case, at least some sort of link between the first-dimensional notion and apriority has been claimed. In Kaplan's and Stalnaker's original publications, it is held that character and diagonal propositions closely reflect matters of apriority, at least in some cases. For Evans and Davies and Humberstone, when a statement of a certain sort is knowable a priori, it is deeply necessary, or true fixedly actually. And for Chalmers and Jackson, whenever a sentence is a priori, it has a necessary primary intension or 1-intension.

But these similarities mask deep underlying conceptual differences. These systems are defined in quite different ways, and apply to quite different items of language, yielding quite different results. Correspondingly, proponents of these systems differ greatly in the scope and strength of their claims. Kaplan's analysis is restricted to just a few linguistic expressions: indexicals and demonstratives. He explicitly resists an extension of his system to other expressions, such as names and natural kind terms. Evans and Davies and Humberstone develop their analysis for a different narrow class of expressions: descriptive names, and perhaps (in the case of Davies and Humberstone) some natural kind terms. Stalnaker's analysis applies in principle to any sentence, but in more recent work, he has explicitly disavowed any strong connection with apriority, and has been skeptical about applications of two-dimensional semantics in that direction. By contrast, Chalmers and Jackson suggest that their notions are defined for a very wide class of expressions, and make strong claims about the connection between these notions and apriority. (The current paper might be viewed in part as a defense of these strong claims.)

These differences arise from different *interpretations* of the formal two-dimensional framework. The framework of worlds and intensions, taken alone, is simply an abstract structure in need of content. Different interpretations flesh out this content in different ways. The interpretations are not necessarily incompatible, although it is possible that some are ill-defined, or rest on false presuppositions. The relations between these interpretations, however, are not well-understood.

The main project of this paper is to explore the different ways in which a two-dimensional framework can be understood. What are the fundamental concepts underlying different interpretations of the framework? How are these related? How do the differences between these interpretations explain the differences in the scope and strength of the claims that are made for them? Which interpretations of the framework yield the strongest connections between the first dimension and the rational domain?

1.4 The Core Thesis

The central question on which I will focus is the following. Is there an interpretation of the two-dimensional framework that yields constitutive connections between meaning, reason, and modality? That is, is there an interpretation on which the first dimension is tied universally to the rational domain? On this way of thinking, the ideal form of the two-dimensional framework will recapture something like the neo-Fregean thesis: two terms will have the same 1-intension if and only if they are equivalent a priori. To get at this question, we can focus on the following core thesis:

> **Core Thesis**: For any sentence S, S is a priori iff S has a necessary 1-intension.

Here, S should be understood as a sentence token (such as an utterance) rather than a sentence type, to accommodate the possibility that different tokens of the same expression type may have different 1-intensions. Correspondingly, we should understand apriority as a property of sentence tokens. I will say more about the relevant notion of apriority and the type/token distinction in Section 3.8. But for now, to a first approximation, we can say that a sentence token S is a priori when S expresses actual or potential a priori knowledge (for the subject who utters S). And I will take it that the intuitive judgments about apriority above are correct: a typical utterance of 'Hesperus is Phosphorus' is not a priori, in this sense, while a typical utterance of 'All bachelors are unmarried' is a priori in this sense.

The Core Thesis links the rational notion of apriority, the modal notion of necessity, and the semantic notion of intension. If the Core Thesis is true, it restores a golden triangle of connections between meaning, reason, and possibility. It also immediately entails a version of the Neo-Fregean Thesis (given plausible principles about compositionality).

> **Neo-Fregean Thesis (2D Version)**: Two expressions 'A' and 'B' have the same 1-intension iff 'A $\equiv$ B' is a priori.

If the two-dimensional framework can be understood in such a way that the Core Thesis is true, it promises an account of a broadly Fregean aspect of meaning, tied constitutively to the epistemic domain. It also promises further rewards: perhaps an account of the contents of thought on which content is tied deeply to a thought's rational role (potentially yielding an account of so-called "narrow content" and "modes of presentation" in thought), and perhaps a view of modality on which there are deep links between the rational and modal domains (potentially grounding a connection between notions of conceivability and possibility). So the key question in what follows will be: can we define 1-intensions so that the Core Thesis is true?

To anticipate, my answer will be as follows. There are two quite different ways of understanding the two-dimensional framework: the *contextual* understanding and the *epistemic* understanding. The contextual understanding uses the first dimension to capture *context-dependence*. The epistemic understanding uses the first dimension to capture *epistemic dependence*. The contextual understanding is more familiar, but it cannot satisfy the Core Thesis. The epistemic understanding is less familiar, but it

can satisfy the Core Thesis. The reason is that only on the epistemic understanding is the first dimension constitutively tied to the epistemic domain.

Within each of these general understandings of the framework, there are various possible specific interpretations. In what follows, I will first explore contextual interpretations (Section 2), and then epistemic interpretations (Section 3). Some of these interpretations are closely related to existing proposals, but rather than working directly with existing proposals, I will characterize these interpretations from first principles. This allows us to examine the properties of these interpretations in a clear light, free of problems of textual exegesis. Later in the paper, I will examine how existing proposals fit into this scheme.

A methodological note: in this paper I will adopt the approach of *semantic pluralism*, according to which expressions can be associated with semantic values in many different ways. Expression types and expression tokens can be associated (via different semantic relations) with extensions, various different sorts of intensions, and with many other entities: structured propositions, conventionally implied contents, and so on. On this approach, there is no claim that any given semantic value exhausts the meaning of an expression, and I will not claim that the semantic values that I focus on are exhaustive. I think that such claims are almost always implausible.

Likewise, this approach gives little weight to disputes over whether a given (purported) semantic value is "the" meaning of an expression, or even whether it is truly a "semantic" value at all. Such disputes will be largely terminological, depending on the criteria one takes to be crucial in one's prior notion of "meaning" or "semantics". On the pluralist approach, the substantive questions are: can expressions (whether types or tokens) be associated with values that have such-and-such properties? If so, what is the nature of the association and of the values? What aspects of language and thought can this association help us to analyze and explain?

My focus in this paper will be almost wholly on whether there is an association between expression tokens and 1-intensions that satisfies the Core Thesis, and on how this association can be understood. I will not say much more about the motivations for this sort of approach, about the broader shape of the resulting semantic theory, or about applications. Motivation and broader questions are discussed in "On Sense and Intension" (Chalmers 2002b), which gives a gentler introduction to these issues. Applications are discussed in "The Components of Content" (Chalmers 2002c) and in "Does Conceivability Entail Possibility?" (Chalmers 2002a).

2. The Contextual Understanding

On the contextual understanding of two-dimensional semantics, the possibilities involved in the first dimension represent possible *contexts of utterance*, and the intension involved in the first dependence represents the *context-dependence* of an expression's extension. There are many ways in which the extension of an expression can depend on the context in which it is uttered. On the contextual understanding, a 1-intension captures the way in which an expression's extension depends on its context. As we will see, this sort of context-dependence can itself be understood in a number of different ways.

To formalize this, we can start by focusing on expression *tokens*: spoken or written tokens of words, sentences, and other expressions. We can take it that any expression token has an extension. In cases where a token "aims" to have an extension but fails, as with an empty name, we can say that it has a null extension. If there are some expression tokens that do not even aim to have an extension (as perhaps with some exclamations), they are outside the scope of our discussion. A token of a sentence corresponds to an utterance; its extension is a truth-value.

Any expression token falls under a number of different expression *types*. A token may fall under an orthographic type (corresponding to its form), a semantic type (corresponding to its meaning), a linguistic type (corresponding to its identity within a language), and various other types. Different tokens of the same expression type will often have different extensions. When two tokens of the same expression type have different extensions, this reflects a difference in the *context* in which the tokens are embedded.

For our purposes, contexts can be modeled as centered worlds. The context in which an expression token is uttered will be a centered world containing the token. This can be modeled as a world centered on the speaker making the utterance, at the time of utterance. It is also possible to model a context by a different sort of centered world with just an expression token marked at the center. The previous version will work for most purposes, however, as long as we assume that a subject makes at most one utterance at a given time.

One can now define the *contextual intension* of an expression type. This is a function from centered worlds to extensions. It is defined at worlds centered on a subject uttering a token of the expression type. At such a world, the contextual intension returns the extension of the expression token at the center.

One can also define the contextual intension of an expression token, relative to a type of which it is a token. This is also a function from centered worlds to extensions. It is defined at worlds centered on a token of the same type, and returns the extension of the token at the center. This contextual intension is the same as the contextual intension of the relevant expression type.

The first-dimensional intensions in the two-dimensional framework are often understood as contextual intensions of some sort. On this way of seeing things, a 1-intension mirrors the evaluation of certain metalinguistic subjunctive conditionals: if a token of the relevant type were uttered in the relevant context, what would its extension be? Of course, for every different way of classing expression tokens under types, there will be a different sort of contextual intension. In what follows I examine some of the relevant varieties of contextual intension.[4]

[4] Constructs akin to contextual intensions have been stressed by Robert Stalnaker in a number of writings (e.g. Stalnaker 1978, 1999). At the same time, Stalnaker and Ned Block have both been active critics of the overextension of this framework (e.g. Stalnaker 1990, 2001; Block 1991; Block and Stalnaker 1999). The discussion in this section owes a significant debt to Stalnaker and Block. Although I carve up the territory in a different way, a number of the varieties of contextual intension that I mention are touched on explicitly or implicitly by Stalnaker and Block at various points, and some of my critical points echo points made by them in criticizing certain applications of the two-dimensional framework.

2.1 Orthographic contextual intensions

We can say that two tokens are tokens of the same *orthographic type* when they have the same orthography. This holds roughly when they are made up of the same letters or sounds, regardless of their meaning, and regardless of the language in which they are uttered. The exact details of what counts as the same orthography can be understood in different ways, but these differences will not matter for our purposes.

The *orthographic contextual intension* of an expression token T is defined at centered worlds with a token of T's orthographic type at the center. It maps such a world to the extension of the relevant token in that world.

(The orthographic contextual intension of a sentence token is closely related to its *diagonal proposition*, as defined by Stalnaker (1978). I will return to this matter later.)

As an example, let S be Oscar's utterance of 'Water is H_2O'. Let W_1 be Oscar's world (Earth), centered on Oscar making this utterance. Oscar's utterance is true, so S's orthographic contextual intension is true at W_1. Let W_2 be a universe containing Twin Earth (where everything is just as on earth except that the watery liquid is XYZ), centered on Twin Oscar uttering 'Water is H_2O'. Twin Oscar's utterance is false (his word 'water' refers to XYZ), so S's orthographic contextual intension is false at W_2. Let W_3 be a universe containing Steel Earth, where the word 'water' refers to steel but chemical terms are the same, centered on Steel Oscar uttering 'Water is H_2O'. Steel Oscar's utterance is false, so S's orthographic contextual intension is false at W_3.

It is clear that orthographic contextual intensions do not satisfy the Core Thesis. For *every* orthographic type, some possible token of that type expresses a falsehood. For example, there are worlds in which the string 'bachelors are unmarried' means that horses are cows. In such a centered world, the orthographic contextual intension of 'bachelors are unmarried' is false. The same goes for any sentence. So no truth has a necessary contextual intension, and in particular no a priori truth has a necessary contextual intension. So if 1-intensions are understood as orthographic contextual intensions, the Core Thesis is obviously false.

2.2 Linguistic contextual intensions

We can say that two expression tokens are tokens of the same *linguistic type* when they are tokens of the same expression in a language. This assumes that expression tokens belong to languages, and that languages involve expressions such as words, phrases, and sentences. So any two tokens of the English word 'water' share a linguistic type, as do any two utterances of the French sentence 'C'est la vie'.

The *linguistic contextual intension* of an expression token T is defined at centered worlds with a token of T's linguistic type at the center. It maps such a world to the extension of the relevant token in that world.

(The linguistic contextual intension of an expression is in some respects like its *character*, as defined by Kaplan. I will return to this matter later.)

As before, let S be Oscar's utterance of 'Water is H_2O'. If W_1 is Oscar's own centered world (Earth): S's linguistic contextual intension is true at W_1. If W_2 is Twin Oscar's centered world (Twin Earth): it is arguable that Twin Oscar's word

'water' is a *different* word from Oscar's word 'water'. Certainly *if* the referent of 'water' is essential to the word, as many theorists hold, then Twin Oscar's 'water' is a different word. If so, S's linguistic contextual intension is not defined at W_2. If W_3 is Steel Oscar's centered world (where 'water' means steel): here it is reasonably clear that Steel Oscar's 'water' is a different word that has the same orthography. If so, S's linguistic contextual intension is not defined at W_3. Applying this sort of reasoning, one reaches the conclusion that S's contextual intension is true at every world in which at which it is defined, since the English word 'water' refers to H_2O in every world in which it exists, and so does the English expression 'H_2O'.

If this is right, then linguistic contextual intensions do not satisfy the core thesis. 'Water is H_2O' is a posteriori, but it seems to have a necessary contextual intension, true at every world at which it is defined. The same goes even more clearly for sentences involving names, such as 'Cicero is Tully'. It is widely held that names have their referents essentially; if so, the linguistic contextual intensions of true identities of this sort will be true at all worlds at which they are defined. As such, linguistic contextual intensions do not behave at all like Fregean senses. If 1-intensions are understood as linguistic contextual intensions, the Core Thesis is false.

There are some expressions for which linguistic contextual intensions behave more like Fregean senses. One such is 'I': setting certain odd cases aside, any token of the English word 'I' picks out the utterer of that token. So the linguistic contextual intension of 'I' picks out the speaker at the center of any centered world at which it is defined. In this way, it behaves much as we earlier suggested the 1-intension of 'I' should behave. Something similar applies to other indexicals, such as 'today', and to some broadly descriptive terms, such as 'philosopher'. It is in the case of names and natural kind terms that the fit seems to be worst.

2.3 Semantic contextual intensions

We can say that two expression tokens are tokens of the same *semantic type* when they have the same semantic value. An expression token's semantic value is its meaning or content, or some aspect of its meaning or content. There are many different ways of assigning semantic values to expression tokens, so there are correspondingly many different ways of classing expression tokens under semantic types.

The *semantic contextual intension* of an expression token T is defined at centered worlds with a token of T's linguistic type at the center. It maps such a world to the extension of the relevant token in that world.

As before, let S be Oscar's utterance of 'Water is H_2O'. If W_1 is Oscar's own centered world (Earth): S's semantic contextual intension is true at W_1. If W_2 is Twin Oscar's centered world (Twin Earth): at least on many ways of assigning semantic value, Twin Oscar's term 'water' has a different semantic value from Oscar's, so S's semantic contextual intension (for this sort of semantic type) is undefined at W_2. If W_3 is Steel Oscar's centered world, then Steel Oscar's term 'water' clearly has a different semantic value from Oscar's, so S's semantic contextual intension is undefined at W_3. If W_4 is a world centered on French Oscar, a counterpart of Oscar who speaks French and is uttering 'eau est H_2O': then it is plausible that this

utterance has the same semantic value as Oscar's, so S's semantic contextual intension is defined at W_1 and is true there.

Of course the behavior of a semantic contextual intension will depend on our choice of semantic value. For example, if we stipulate that the relevant semantic value of an expression is its *extension*, then any two co-extensive expressions will have the same semantic contextual intension, and there is no chance that the Core Thesis will be true. There are two choices of semantic value that are somewhat more interesting, however.

We might stipulate that the relevant semantic value of an expression is its *standing meaning*: roughly, the aspect of meaning that is common to all tokens of the expression's linguistic type. If we do this, then an expression's semantic contextual intension will be an extension of its linguistic contextual intension to a broader space of worlds. At worlds centered on a token of the same linguistic type, the intensions will give the same results. But the semantic contextual intensions will also be defined at other worlds, centered on synonyms and translations of the original expression. Nevertheless, if 1-intensions are understood as these semantic contextual intensions, the Core Thesis will be false for the same reasons as in the case of linguistic contextual intension. For example, if the extension of 'water' is essential to the word, then it is part of the word's standing meaning. So the semantic contextual of 'Water is H_2O' will be true at every world where it is defined, and the Core Thesis is false.

Alternatively, we might stipulate that the relevant semantic value of an expression token is its *Fregean or descriptive content*, corresponding roughly to the expression's cognitive significance for the subject. On this reading, the Core Thesis may be more plausible. For example, one might argue that Oscar's and Twin Oscar's terms 'water' have the same descriptive content. If so, then the semantic contextual intension of Oscar's utterance 'Water is H_2O' is defined at W_2 and is false there. On the other hand, Steel Oscar's term 'water' plausibly has a different descriptive content, so the semantic contextual intension of Oscar's utterance is not defined at W_3.

Understood this way, semantic contextual intensions behave as we might expect a Fregean 1-intension to behave, at least to some extent. One can argue that when a statement is a priori, any possible statement with the same descriptive content will be a priori and so will be true, so that the expression's semantic contextual intension will be necessary, as the Core Thesis requires. Correspondingly, one might suggest that when a statement is not a priori, then there will be possible statements with the same descriptive content that are false, so that the statement's semantic contextual intension will not be necessary, as the Core Thesis requires.

I will argue shortly that this is not quite right. But even if it were right, it is clear that this sort of 1-intension cannot underwrite the full ambitions of the Fregean two-dimensionalists. The Fregean two-dimensionalist, as sketched previously, intends to use the two-dimensional framework to *ground* an aspect of meaning that is constitutively tied to meaning. But semantic contextual intensions as defined here *presuppose* such a Fregean semantic value, and so cannot independently ground such an account. If this is the best a two-dimensionalist can do, then if someone is independently doubtful about a Fregean aspect of meaning, two-dimensionalism

cannot help. At best, two-dimensionalism will be a helpful tool in analyzing such a notion of meaning, given an independent grounding for the notion.[5]

2.4 A further problem

We have seen that orthographic contextual intensions are far from satisfying the Core Thesis, while linguistic contextual intensions are closer at least in some cases, and some sort of semantic contextual intensions may be closer still. But there is a further problem that arises for any sort of linguistic or semantic contextual intension, suggesting that no such contextual intension can satisfy the Core Thesis.

Let S be a token of 'A sentence token exists' (where a sentence token is understood to be a concrete entity produced by speech, writing, or a similar process). Then S is true. Furthermore, any token of the linguistic item 'A sentence token exists' is true. Any token that *means* the same thing as 'A sentence token exists' is true. So it seems that S will have a necessary linguistic contextual intension, and a necessary semantic contextual intension, under any reasonable way of classifying linguistic and semantic types. But S is clearly a posteriori: it expresses empirical knowledge of the world, which could not be justified independently of experience. So S is a counterexample to the Core Thesis. So the Core Thesis is false for any sort of semantic or linguistic contextual intension.

The same goes for a number of other sentences. If S_1 is 'Language exists' (where a language is understood to be a spoken or written language, not just an abstract language), then any utterance of the same expression or with the same meaning will be true. So S_1 has a necessary linguistic and contextual intension, despite being a posteriori. If S_2 is 'I exist', then any utterance of the same expression with the same meaning will be true, so S_2 has a necessary linguistic and semantic contextual intension. But (somewhat controversially) S_2 is a posteriori, justifiable only on the basis of experience. If S_3 is 'I am uttering now', then any utterance of the same expression or with the same meaning will be true. S_3 is clearly a posteriori, but has a necessary linguistic and semantic contextual intension.

All these cases are counterexamples to the Core Thesis. All of them are a posteriori and cognitively significant, and many of them seem to be as cognitively significant as paradigmatic expressions of empirical knowledge. But all have necessary semantic and linguistic contextual intensions. So the Core Thesis is false for all such intensions.

The trouble is that apriority and being true whenever uttered are fundamentally different notions. The first builds in an epistemic or rational element, but the second builds in no such element. The second notion builds in a metalinguistic element, but the first builds in no such element. It is possible to understand the second in a way that makes it coincide with the first in many cases, in effect by building in an epistemic element into the individuation of the relevant linguistic types. But it is impossible to do so in all such cases, since the second has an ineliminable metalinguistic element that goes beyond the epistemic domain.

[5] This sort of point is made quite clearly, in the context of discussing narrow content, by Stalnaker (1991), Block (1991), and Block and Stalnaker (1999).

I think the moral is that to satisfy the Core Thesis, we must understand the two-dimensional framework in a quite different, non-contextual way. But before doing so, I will more briefly examine some further ways in which one might define a contextual intension.

2.5 Hybrid contextual intensions

Given orthographic, linguistic, and semantic types for expression tokens, it is possible to define *hybrid types* corresponding to conjunctions of two or more of these types. One can then define corresponding *hybrid contextual intensions*.

For example, one might say that two expressions share the same *orthographic/semantic type* when they share the same orthographic type and the same semantic type. One can then define the *orthographic/semantic contextual intension* of an expression as the function that maps a world centered on a token of the appropriate orthographic/semantic type to the extension of that token.

Hybrid contextual intensions may be useful for some purposes, but it is clear that they will not satisfy the Core Thesis any better than non-hybrid contextual intensions. So I will set them aside here.

2.6 Token-reflexive contextual intensions

It is possible to define a slightly different sort of contextual intension for an expression token by focusing not on the types that the token falls under, but on the token itself. Let us assume that expression tokens are not tied to their context essentially: a given token might have been uttered in another context. Then we can say that the *token-reflexive* contextual intension of an expression token T is a function that maps a centered world containing T to the extension of T in that world.

The precise behavior of a token-reflexive contextual intension will depend on what properties an expression token has necessarily. It is plausible that if such a token has any properties necessarily, it has its orthographic properties necessarily. If so, its token-reflexive contextual intension will be a restriction of its orthographic contextual intension, obtained by eliminating worlds centered on a *different* token of the same orthographic type. One might also hold that a token has some semantic value necessarily, or that it has its linguistic type necessarily. If so, its token-reflexive contextual intension will be a restriction of its semantic or linguistic contextual intension. If an expression has more than one of these things necessarily, its token-reflexive contextual intension will be a restriction of a hybrid contextual intension. If it has further properties necessarily (e.g. its speaker), it will be a further restriction of the relevant contextual intension.

It is not obvious how to decide exactly which properties an expression token has necessarily. But however we do this, it is clear that token-reflexive contextual intensions cannot satisfy the Core Thesis. The counterexamples discussed above, such as 'I am uttering now', will apply equally to token-reflexive contextual intensions. Furthermore: insofar as tokens have any properties necessarily, one can likely construct sentence tokens attributing these properties that are true whenever uttered, but not a priori (e.g. 'This token has four words'; 'David Chalmers is speaking now'). And

insofar as tokens have few properties necessarily, one can likely construct sentences that are a priori but that are not true whenever uttered (e.g. 'All bachelors are unmarried'). So if 1-intensions are understood as token-reflexive contextual intensions, the Core Thesis is false.

2.7 Extended contextual intensions

In an attempt to get around the problems posed by sentences such as 'I am uttering now', one might attempt to construct contextual intensions that are defined at centered worlds that do not contain a token of the relevant expression type. The most obvious way to do this is to appeal to certain counterfactual conditionals. Let us say that the *extended contextual intension* is defined at any centered world, independently of whether a token of the type is present there. At a given centered world, the extended contextual intension returns what the extension of a token of that type *would be*, if it *were* uttered at the center of that world.

One can then say that the extended contextual intension of an expression token (relative to a type) maps a centered world to what the extension of a token of that type would be, if it were uttered at the center of the world. So in principle, one might have extended linguistic contextual intensions, extended semantic contextual intensions, and so on. One could define an extended token-reflexive contextual intension in an analogous way.

An obvious problem here is that in many cases, it is unclear how to evaluate the counterfactual. It may be reasonably straightforward in some cases, such as 'I am a philosopher': true just when an utterance of 'I am a philosopher' by the subject at the center would be true, so true just when the person at the center is a philosopher. But how is one to evaluate what a token of 'water' would refer to if it were used in a world where there is no liquid, and in which nobody speaks a language? How does one evaluate whether an utterance of 'I am speaking loudly' would be true if it were uttered, in a world where the subject at the center is not in fact speaking? In some cases, it seems impossible for a token of the relevant type to be uttered in the relevant context. In other cases, it may be possible, but it is possible in many different ways, yielding many different results. So the truth of the relevant counterfactuals seems to be underdetermined, and an expression's extended contextual intensions seems to be ill-defined.[6]

Another problem: even if extended contextual intensions behave coherently, they give results that are different from what we need. For example, let S = 'I am uttering now'. S is a posteriori, so the Core Thesis requires that its 1-intension be false at some worlds. For example, it is desirable that S's 1-intension be false at an utterance-free world. Let W be such an utterance-free centered world. To evaluate S's contextual intension at W, we ask: if S were uttered at the center of W, what would its extension be? It is clear that if S *were* uttered in W, it would be true. So S's extended contextual intension is true at W, and indeed is true at all worlds. So the Core Thesis is still false for extended contextual intensions.

[6] A point of this sort is made by Stalnaker (1990).

To get anything like the result that is needed, we would need to evaluate S's extension in W *without* S being present in W. But it is very hard to do that on the contextual model. On the contextual understanding, 1-intensions are derivative on facts about the extensions of various possible tokens, as uttered in various possible contexts. It seems clear that on such an understanding, the 1-intension of a sentence such as 'There are sentence tokens' will never be false.

I think that the idea of an extended contextual intension is getting at something important: that we need to be able to evaluate an expression's 1-intension in centered worlds that do not contain a token of the expression. But this is the wrong way to achieve the goal. To do this properly, I think we need to go beyond the contextual understanding of 1-intensions.

2.8 Cognitive contextual intensions

One might suggest that to capture a token's cognitive significance, we should not focus on a token's broadly *linguistic* properties, such as its orthography, its semantic value, and its language. Instead, we need to focus on its *cognitive* properties, which correspond to mental features of the subject that produces the token. Some such features include: the concept or belief that the token expresses; the cognitive role associated with the token; and the intentions associated with the token. Assuming that we have a way of individuating the mental types in question, we can then classify expression tokens under corresponding cognitive types.

For a given scheme of cognitive typing, one can then define the *cognitive contextual intension* of an expression token as the intension that maps a world centered on a token of the same cognitive type to the extension of that token. In the three cases above: a *conceptual* contextual intension will be defined at worlds centered on a token expressing the same concept or belief; a *cognitive-role* contextual intension will be defined at worlds centered on a token associated with the same cognitive role; and an *intention-based* contextual intension will be defined at worlds centered on a token associated with the same intentions.

Assuming that one can make sense of the relevant typing, there is a natural extension of this idea. One could define a sort of *extended* cognitive contextual intension, defined at worlds that do not contain the token at all, but merely contain the relevant mental feature. For example, the extended conceptual contextual intension will be defined at any world that contains the relevant *concept* at its center, irrespective of whether it contains any token, and will return the extension of the concept. (This assumes that concepts have extensions, which seems reasonable enough.) The extended cognitive-role contextual intension might be defined at any world centered on a concept that plays the relevant cognitive role, returning the concept's extension; and the extended intention-based contextual intension will be defined at any world centered on a concept that is associated with the same intentions.

This sort of intension has some promise of dealing with the central problems raised so far. In the case of 'A sentence token exists': one can make a case that the extended conceptual contextual intension of this expression is *false* at some centered worlds: those in which a subject has the relevant concepts and the relevant thought, but in

which there are no sentence tokens. So the intension is not necessary, reflecting the aposteriority of the sentence. The same goes for 'Language exists', and for 'I am uttering now'. By allowing intensions to be evaluated without relying on language, the metalinguistic element of contextual intensions has been reduced or eliminated.

Still, analogous problems arise. 'I am thinking now' will plausibly have a necessary conceptual contextual intension, but it is plausibly a posteriori: the thought itself is justified only by experience, albeit by introspective experience. The same goes for 'I exist'. And the same will apply to specific attributions of mental features: a thought such as 'I have the concept *concept*' will be true whenever it is thought, but it is not justifiable a priori. Something similar applies to thoughts attributing certain cognitive roles or certain intentions. So even here, some a posteriori sentences and thoughts will have a necessary 1-intension.

As for the other main sort of problem discussed so far, that associated with 'Water is H_2O': a proponent might hold that although Oscar and Twin Oscar do not have the same *word* 'water', their words express the same *concept*, at least under one reasonable way of individuating concept types. If so, then the conceptual contextual intension of Oscar's token 'Water is H_2O' will be false at the world centered on Twin Oscar, as the Fregean conception requires. At the same time, it might be undefined at the world centered on Steel Oscar (since he seems to have a different concept), as required.

It is controversial, however, whether concepts (or roles or intensions) can be individuated in a way that yields these results. Many theorists hold that even a token concept expressed by 'water' has its extension essentially, and that all concepts of the same type have the same extension. If so, then a statement such as 'Water is H_2O' will have a necessary intension. They might concede that concepts or thoughts can also be individuated syntactically or formally; but on this way of doing things, 'All bachelors are unmarried' will have a contingent intension. So either way, the Core Thesis is false.

One might argue that there is an intermediate way of individuating concept types that yields the right results. But many will deny this. It might be objected that this requires individuating concepts by their *narrow content* (that aspect of their content that is determined by a subject's intrinsic properties), and it is highly controversial whether narrow content exists. Some think that the two-dimensional framework can be used to give an account of narrow content; but in this context, it seems illegitimate for the framework to presuppose narrow content. This is a precise analog of the problem that arose for the Fregean version of semantic contextual intensions above.

I think that the situation here is not entirely clear. One could argue with some plausibility that there is an *intuitive* sense in which Oscar and Twin Oscar have the same concept, where there is no corresponding intuitive sense that they have the same word. If so, one could appeal to this intuitive sort of concept individuation to ground some sort of conceptual contextual intension here. One might arguably be able to do the same sort of thing with cognitive roles, or intentions. But the intuitions in question are likely to be disputed by many, so this approach will be at best weakly grounded, unless one can give some sort of independent account of the relevant concept types.

On my view, (extended) cognitive contextual intensions are the sort of contextual intensions that are closest to satisfying the Core Thesis. But ultimately, the central problems arise for them too. One might try appealing to related notions that carry features of the subject across worlds: for example, an *evidential* contextual intension, requiring sameness of evidence; a *fixing* contextual intension, requiring sameness of reference-fixing procedures or intentions; a *physical* contextual intension, requiring that subjects be physical duplicates; *functional*, *phenomenal*, *physical-phenomenal* contextual intensions, which require that subjects be functional, phenomenal, and physical-phenomenal duplicates; and so on. But it is not hard to see that all of these suggestions are subject to versions of the problems mentioned above. So we still need an account of the relevant intensions.

2.9 Summary

Overall, it seems that there is no way to define contextual intensions so that they satisfy the Core Thesis. Two central problems have arisen repeatedly. First, by building in a token of the relevant mental or linguistic type into the world of evaluation, the constitutive connection with the a priori is lost. Second, for a contextual intension to behave in a quasi-Fregean manner, we need to antecedently classify tokens under some sort of quasi-Fregean type, so that the framework cannot independently ground quasi-Fregean notions, as was originally hoped.

Contextual intensions may still be useful for many purposes. But they do not yield any restoration of the golden triangle, and in particular they do not deliver a notion of meaning that is deeply tied to reason. The fundamental problem is that although some contextual intensions yield a reasonably strong correlation with the epistemic domain, none is *constitutively* connected to the epistemic domain. To restore the connection between meaning and reason, we need to approach the two-dimensional framework in epistemic terms.

3. The Epistemic Understanding

3.1 Epistemic dependence

On the epistemic understanding of two-dimensional semantics, the possibilities involved in the first dimension are understood as *epistemic possibilities*, and the intensions involved in the first dimension represent the *epistemic dependence* of the extension of our expressions on the state of the world.

There are two key ideas here. The first is the idea of *epistemic space*: there are many ways the world might turn out to be, and there is a corresponding space of epistemic possibilities. The second is the idea of *scrutability*: once we know how the world has turned out, or once we know which epistemic possibility is actual, we are in a position to determine the extensions of our expressions. Together, these two ideas suggest that an expression can be associated with a function from epistemic possibilities to extensions: an *epistemic intension*.

Take the first idea first. There are many ways the world might be, for all we know. And there are even more ways the world might be, for all we know a priori. The

oceans might contain H_2O or they might contain XYZ; the evening star might be identical to the morning star or it might not. These ways the world might be correspond to epistemically possible hypotheses, in a broad sense. Let us say that a claim is *epistemically possible* (in the broad sense) when it is not ruled out a priori. Then it is epistemically possible that water is H_2O, and it is epistemically possible that water is XYZ. It is epistemically possible that Hesperus is Phosphorus, and epistemically possible that Hesperus is not Phosphorus.

Just as one can think of metaphysically possible hypotheses as corresponding to an overarching space of metaphysical possibilities, one can think of epistemically possible hypotheses as corresponding to an overarching space of epistemic possibilities. Some possibilities in the space of metaphysical possibilities are maximally specific: these can be thought of as *maximal metaphysical possibilities*, or as they are often known, possible worlds. In a similar way, some possibilities in the space of epistemic possibilities are maximally specific: these can be thought of as *maximal epistemic possibilities*, or as I will call them, *scenarios*.

A scenario corresponds, intuitively, to a maximally specific way the world might be, for all one can know a priori. Scenarios stand to epistemic possibility as possible worlds stand to metaphysical possibility. Indeed, it is natural to think of a scenario as a sort of possible world, or better, as a *centered* possible world. There are some complications here, but for the moment it is helpful to think of scenarios intuitively in such terms.

For any scenario, it is epistemically possible that the scenario is actual. Intuitively speaking, for any qualitatively specified centered world W, it is epistemically possible that W is actual. Here the center represents a hypothesis about my own location within the world. In entertaining the hypothesis that W is actual, I entertain the hypothesis that the actual world is qualitatively just like W, that I am the subject at the center of W, and that now is the time at the center of W.

For example, let the XYZ-world be a specific centered "Twin Earth" world, in which the subject at the center is surrounded by XYZ in the oceans and lakes. Then no amount of a priori reasoning can rule out the hypothesis that the XYZ-world is my actual world: i.e., that I am in fact living in such a world, where the liquid in the oceans and lakes around me is XYZ. So the XYZ-world represents a highly specific epistemic possibility.

When we think of a world as an epistemic possibility in this way, we are *considering it as actual*. On the epistemic understanding, to consider a world W as actual is to consider the hypothesis that W is one's own world. When one considers such a hypothesis, in effect one considers the hypothesis that D is the case, where D is a statement giving an appropriate description of W. One can think of D, intuitively, as a description of W in neutral qualitative terms, along with a specification in indexical terms of a center's location in W. I will return to this matter later.

The second key idea is that of scrutability: the idea that there is a strong epistemic dependence of an expression's extension on the state of the world. If we come to know that the world has a certain character, we are in a position to conclude that the expression has a certain extension. And if we were to learn that the world has a different character, we would be in a position to conclude the expression has a different

extension. That is: we are in a position to come to know the extension of an expression, depending on which epistemic possibility turns out to be actual.

If we take the case of 'Water is H_2O': we can say that given that the world turns out as it actually has, with H_2O in the oceans and lakes, then it turns out that water is H_2O. So if the H_2O-world is actual, water is H_2O. But if we were to discover that the oceans and lakes in the actual world contained XYZ, we would judge that water is XYZ. And even now, we can judge: *if* it turns out that the liquid in the oceans and lakes is XYZ, it will turn out that water is XYZ. Or we can simply say: if the XYZ-world is actual, then water is XYZ.

The same goes more generally. If W_1 is a specific scenario in which the morning and evening stars are the same, and W_2 is a scenario in which the morning and evening stars are different, then we can say: if W_1 is actual, then Hesperus is Phosphorus; if W_2 is actual, then Hesperus is not Phosphorus. The same goes, in principle, for a very wide range of scenarios and statements. Given a statement S, and given enough information about an epistemically possible state of the world, we are in a position to judge whether, *if* that state of the world obtains, S is the case.

All this is reflected in the way we use language to describe and evaluate epistemic possibilities. It is epistemically possible that water is XYZ. It is also epistemically possible that the XYZ-world is actual. And intuitively speaking, the epistemic possibility that the XYZ-world is actual is an *instance* of the epistemic possibility that water is XYZ. We can say as above: *if* the XYZ-world turns out to be actual, it will turn out that water is XYZ. We might also use a straightforward indicative conditional: if the XYZ-world is actual, then water is XYZ. Or we can use the Ramsey test, commonly used to evaluate indicative conditionals: if I hypothetically accept that the XYZ-world is actual, I should hypothetically conclude that water is XYZ.

We can put all this by saying that the XYZ-world *verifies* 'Water is XYZ', where verification is a way of expressing the intuitive relation between scenarios and sentences described above.[7] Intuitively, a scenario W verifies a sentence S when the epistemic possibility that W is actual is an instance of the epistemic possibility that S is the case; or when we judge that if W turns out to be actual, it will turn out that S is the case; or if the indicative conditional 'if W is actual, then S is the case' is rationally assertible, or if hypothetically accepting that W is actual leads to hypothetically concluding that S is the case. We can also say that when W verifies S, W makes S true when it is *considered as actual*. Verification captures the way that we use language to describe and evaluate epistemic possibilities.

This dependence can be represented by the *epistemic intension* of a sentence S. This is a function from scenarios to truth-values. If a scenario W verifies S, then S's epistemic intension is true at W; if W verifies ~S, then S's epistemic intension is false at W; otherwise, S's epistemic intension is indeterminate at W. So the epistemic intension of 'Water is XYZ' is true at the XYZ-world.

Given this intuitive conception of epistemic intensions, there is a strong prima facie case that they satisfy the Core Thesis. When S is a priori, we would expect that

[7] The term 'verify' is used for a related idea in Evans (1979). See also Yablo (1999).

every scenario verifies S. And when S is not a priori, ~S is epistemically possible, so we would expect that there is a scenario that verifies ~S. If these claims hold true, then S is a priori iff S has a necessary epistemic intension (one that is true at all scenarios).

Epistemic intensions resemble contextual intensions in some superficial respects, but they are fundamentally quite different. The central difference, as we will see, is that epistemic intensions are defined in epistemic terms. From what we have seen so far, epistemic intensions behave at least somewhat as one would like a quasi-Fregean 1-intension to behave. But to investigate this matter, we must define the relevant notions more precisely.

3.2 Epistemic intensions

The intuitive picture of the epistemic understanding above can be regarded as capturing what is essential to an epistemic understanding. To fill in the picture, however, a more precise analysis is required. What follows is one way to flesh out these details. Not all of the details that follow are essential to an epistemic account *per se*, but they provide a natural way of elaborating such an account.

Starting with the intuitive picture, we can say that the epistemic intension of a sentence token is a function from a space of scenarios to the set of truth-values, such that:

> The epistemic intension of a sentence token S is true at a scenario W iff the hypothesis that W is actual epistemically necessitates S.

When the conditions specified here obtain, we can also say that W *verifies* S. The epistemic intension of S will be false at W when W verifies ~S, and it will be indeterminate at W when W verifies neither S nor ~S.

Rather than leaving the notion of "the hypothesis that W is actual" as primitive, it is useful (although not mandatory) to invoke the notion of a *canonical description* of a scenario. We can then characterize an epistemic intension as follows.

> The epistemic intension of a sentence token S is true at a scenario W iff D epistemically necessitates S, where D is a canonical description of W.

It remains to clarify three notions: the notion of a scenario, that of a canonical description, and that of epistemic necessitation. I investigate each of these in what follows.[8]

3.3 Epistemic necessitation

First, we need to say more about epistemic possibility and necessity. The epistemic understanding of two-dimensional semantics is grounded in a notion of *deep epistemic possibility*, or equivalently, of *deep epistemic necessity*. In the ordinary sense, we say

[8] Note that some of these details are necessarily complex, and some readers may prefer to skim the remainder of this section or skip ahead to Section 4 on a first reading. Some other papers cover some of this material in more depth: notably, "The Nature of Epistemic Space" (Chalmers forthcoming), which covers the issues in 3.4 in more detail; "Conceptual Analysis and Reductive Explanation" (Chalmers and Jackson 2001), which is especially relevant to the issues in 3.6; and "Does Conceivability Entail Possibility?" and "On Sense and Intension" (Chalmers 2002a and 2002b), which discuss a number of aspects of these issues that are not discussed here.

that S is epistemically possible roughly when S may be the case for all we know, and that S is epistemically necessary roughly when we are in a position to know that S is the case. A notion of deep epistemic necessity goes beyond this sort of dependence on the shifting state of an individual's knowledge, to capture some sort of rational *must*: a statement is deeply epistemically necessary when in some sense, it rationally must be true.

Such a notion can be understood in various ways, but for our purposes there is a natural candidate. We can say that S is deeply epistemically necessary when it is *a priori*: that is, when the thought expressed by S expresses actual or potential a priori knowledge. (I say more about the notion of apriority in Section 3.9.) Then S is deeply epistemically possible when the negation of S is not epistemically necessary: that is, when the thought that S expresses cannot be ruled out a priori. Henceforth, I will usually drop the modifiers "deep" and "deeply", and speak simply of epistemic possibility and necessity.

In this sense, 'Water is XYZ' is epistemically possible: one cannot know a priori that water is not XYZ. In the same way, 'Hesperus is not Phosphorus' is epistemically possible, as is 'I am not a philosopher'. On the other hand, 'Some bachelors are married' is not epistemically possible, and 'All bachelors are married' is epistemically necessary. Similarly, one can argue that 'Hesperus is not Hesperus' is epistemically impossible, and that its negation is epistemically necessary.

A claim is deeply epistemically possible, intuitively speaking, when it expresses a rationally coherent hypothesis about the actual world. The standards of rational coherence here are in one sense weaker than usual: if a hypothesis conflicts with empirical knowledge, it may still be deeply epistemically possible. The standards are in another sense stronger than usual: if a hypothesis can be ruled out only by a great amount of a priori reasoning, it is nevertheless deeply epistemically impossible. It is possible to define notions of possibility that meet different standards, but the current standards are best for our current purposes.

The epistemic necessity operator applies to both sentence types and sentence tokens. We require this as the sentences S whose epistemic intensions we are defining are tokens, and it is possible for two sentence tokens of the same linguistic type to have different epistemic properties (for the reasons, see Section 3.8). The canonical descriptions D of scenarios, on the other hand, are sentence types, using expressions whose epistemic properties are fixed by the language. We also need an epistemic necessitation operator between sentence types of this sort and sentence tokens.

An epistemic necessity operator of this sort can be seen as a primitive of the system I am developing. On the picture where epistemic necessity corresponds to apriority, we can characterize its properties intuitively as follows. Let us say that *thoughts* are the sort of occurrent propositional attitudes expressed by assertive sentences. Then a sentence token S is epistemically necessary when the thought expressed by S can be justified independently of experience, yielding a priori knowledge. A sentence type D is a priori when it is possible for a token of S to be epistemically necessary. A sentence type D epistemically necessitates a sentence token S when a material conditional '$D \supset S$' is epistemically necessary, where this is understood as a possible token material conditional whose constituent token of S expresses the same thought as the original

token. I will say more about the characterization of epistemic necessity in Section 3.9, but this understanding will suffice for present purposes.

We can now say that a scenario W verifies a sentence token S when a material conditional 'D ⊃ S' is epistemically necessary, where D is a canonical description of W. If epistemic necessity is understood as apriority, then on this model a scenario W verifies a sentence S when one could in principle rule out a priori the hypothesis that W is actual but S is not the case.

This definition works naturally with the characterizations we will give of scenarios and of canonical descriptions, but it should be noted that this is not the only possible definition. There are various ways in which an epistemic framework might characterize the required relationship between D and S in other terms, which need not appeal directly to notions such as apriority.

For example, one might appeal to the intuitive heuristics described earlier. One could say that W verifies S when the epistemic possibility that W is actual is an instance of the epistemic possibility that S is the case. Or appealing to canonical descriptions, one could say that W verifies S when the epistemic possibility that D is the case is an instance of the epistemic possibility that S is the case. Here one might leave this intuitive evaluation of epistemic possibilities as a primitive, much as the intuitive evaluation of counterfactual possibilities is often taken as a primitive.

Alternatively, one could ground epistemic necessitation in indicative conditionals: D epistemically necessitates S when the indicative conditional 'if D is the case, then S is the case' is intuitively acceptable on rational reflection. (See Chalmers (1998) for a discussion of this approach.) In a closely related idea, one could ground epistemic necessitation in the Ramsey test: D epistemically necessitates S (relative to a subject) when if the subject hypothetically accepts that D is the case, the subject should rationally conclude that S is the case. The latter approach yields what we might call the *Ramsey intension* of an expression: the Ramsey intension of a subject's expression S is true at W when if the subject hypothetically accepts that D is the case (where D is a canonical description of W), the subject should rationally conclude that S is the case.

Ramsey intensions behave very much like epistemic intensions as defined above. It is plausible they often yield the same results: for example, both the epistemic intension and the Ramsey intension of "water is H_2O" are plausibly false at the XYZ-world. There are arguably some cases where they yield different results. For example, Yablo (2002) has argued that the indicative conditional "if 'tail' means leg, then tails are legs" is acceptable. If so, then the Ramsey intension of 'tails are legs' may be true in a world where 'tail' means legs, but the epistemic intension will not. (See Chalmers (2002a) for discussion.) Likewise, if I accept that I have recently been given a drug that corrupts my adding abilities, then I should arguably suspend judgment about whether 57 plus 46 is 103. If so, the Ramsey intension of '57 + 46 = 103' will plausibly be indeterminate in a scenario where the subject at the center has been given such a drug, but the epistemic intension will not. It may be that the Ramsey test can be understood in a way that handles the cases above differently, so that Ramsey intensions behave in the way that a Fregean intension should, but the matter is not entirely clear.

Ramsey intensions are a sort of epistemic intension in the general sense, as they are defined in epistemic terms. But where epistemic intensions as defined above are grounded in the notion of apriority, Ramsey intensions are grounded in the notion of rational inference. This has certain advantages: for example, those who are skeptical about apriority usually still accept that there is a coherent notion of rational inference. In what follows I will usually stay with epistemic intensions grounded in a notion of apriority, but the possibility of alternative understandings should be kept in mind.

These alternative understandings suggest that the epistemic understanding of the two-dimensional framework is not entirely beholden to the notion of apriority. Even if one rejects apriority, or if one rejects the application of apriority in this context, one should not reject the epistemic understanding. It is a prima facie datum that there is an epistemic dependence between epistemic possibilities and sentence tokens of the sort that was intuitively characterized earlier. One who rejects apriority will simply need to capture this dependence in other ways. My own view is that the understanding in terms of apriority runs the deepest, but the alternatives deserve exploration.

We can here note a fundamental difference between all of these sorts of epistemic evaluation and contextual evaluation. To evaluate a sentence S in a scenario W, there is no requirement that W contain a token of S. Even if W contains such a token, the definition gives it no special role to play. All that matters is the first-order epistemic relation between D and S, not whether D says something metalinguistic about a token of S. More generally, metalinguistic facts about how a token of S would behave in certain possible circumstances play no role in defining epistemic intensions. This enables us to deal straightforwardly with the problem cases for contextual intensions.

3.4 Scenarios

Scenarios are intended to stand to epistemic possibility as possible worlds stand to metaphysical possibility. This claim can be expressed by the following:

> **Plenitude Principle**: For all S, S is epistemically possible if and only if there is a scenario that verifies S.

In effect, the Plenitude Principle says that there are enough scenarios to verify every epistemically possible claim, and that no scenario verifies an epistemically impossible claim. It is easy to see that if we understand epistemic necessity as apriority, the Plenitude Principle is equivalent to the Core Thesis. (I give it a different name to leave open the option of understanding epistemic necessity in different terms.) So the only question is whether we can understand scenarios and verification so that the Plenitude Principle is true.

Intuitively, a scenario should correspond to a maximally specific epistemically possible hypothesis, or (for short) a maximal hypothesis: a hypothesis such that if one knew that it were true, one would be in a position to know any truth by reasoning alone. (Note that talk of "hypotheses" here is intuitive; formalizations of the relevant notions will follow.) We might say that a hypothesis H1 leaves another hypothesis H2 open if the conjunctions of H1 with both H2 and its negation are epistemically possible. A maximal hypothesis is one that leaves no possible hypothesis open. To every scenario, there should correspond a maximal hypothesis, and vice versa.

3.4.1 Scenarios as centered worlds

There are two concrete ways in which we might understand scenarios. The first is the way we have already sketched: as centered possible worlds. The uncentered part of the world corresponds to a hypothesis about the objective character of one's world. The centered part is needed to handle indexical claims, such as "I am in Australia". If we are given only a full objective description of a world, numerous indexical hypotheses will be left open, so such a description does not correspond to a maximal hypothesis. Correspondingly, there are numerous epistemically possible (but incompatible) objective-indexical claims: for example "the world is objectively thus and I am a philosopher" and "the world is objectively thus and I am not a philosopher". We need distinct scenarios to verify these claims: hence centered worlds.

There is good reason to believe that for every centered world, there is a corresponding maximal hypothesis, at least if we describe worlds under the right sort of canonical description. (It is arguable that for certain indexical hypotheses involving demonstratives, one may need further information in the center of the world: marked experiences, as well as a marked subject and time. But I will leave this matter to one side.) And one can easily make the case that an epistemically impossible sentence will be verified by no centered world (if it were so verified, it would not be epistemically impossible). The residual question is whether there are enough centered worlds to correspond to *all* maximal hypotheses, and to verify all epistemically possible statements. This matters turns on the following thesis:

> **Metaphysical Plenitude**: For all S, if S is epistemically possible, there is a centered metaphysically possible world that verifies S.

The standard Kripkean cases of statements that are epistemically possible but metaphysically impossible are straightforwardly compatible with this thesis. For each such statement S, there is *some* way the world could turn out such that if things turn out that way, it will turn out that S is the case; and each of these ways the world could turn out can be seen as a centered world. In the case of 'Water is XYZ', the XYZ-world is such a world; something similar applies to other cases. One might worry about how a metaphysically possible world (the XYZ-world) can verify a metaphysically impossible statement ('Water is XYZ'). But two-dimensional evaluation makes this straightforward: 'Water is XYZ' is true at the XYZ-world considered as actual, but false at the XYZ-world considered as counterfactual. The metaphysical impossibility of 'Water is XYZ' reflects the fact that it is false at all worlds considered as counterfactual. But this is quite compatible with its being true at some worlds considered as actual.

Are there any counterexamples to the Metaphysical Plenitude thesis? I have argued elsewhere (Chalmers 2002a) that there are no such counterexamples. Certainly, there are no *clear* cases of epistemically possible claims that are verified by no centered world. Still, some controversial philosophical views entail that there are such cases. For example, some theists hold that it is necessary that an omniscient being exists, while also holding that it is not a priori that an omniscient being exists. If so, "No omniscient being exists" will be a counterexample to Metaphysical Plenitude: it will

be an epistemically possible statement that is verified by no possible world. In effect, on this view the space of metaphysical possibilities is *smaller* in some respects than the space of epistemic possibilities.

The same goes for some other philosophical views. On some views on which the laws of nature of our world are the laws of all worlds, for example, the negation of a law of nature will be a counterexample to Metaphysical Plenitude. On views on which a mathematical claim (such as the Continuum Hypothesis) can be necessarily true but not knowable a priori, the negation of such a claim will be a counterexample to Metaphysical Plenitude. On some versions of the epistemic theory of vagueness, some claims involving vague terms (e.g. the statement that someone of a certain height is tall) may be a counterexample to Metaphysical Plenitude. On some materialist views about consciousness, the claim that there are zombies (unconscious physical duplicates of conscious beings) may be a counterexample to Metaphysical Plenitude. If these views are correct, there will be epistemically possible claims that are not verified by any centered metaphysically possible worlds. If so, Metaphysical Plenitude (and the Core Thesis for epistemic intensions over centered metaphysically possible worlds) will be false.

All of these views are highly controversial, and I have argued elsewhere (Chalmers 2002a) that all of them are incorrect. One can plausibly argue in reverse: the Metaphysical Plenitude thesis, which appears to fit all standard cases, gives us reason to reject these controversial views. More deeply, one can argue that these views rest on a mistaken conception of metaphysical possibility and necessity. My own view is that a careful analysis of the roots of our modal concepts supports constitutive links between epistemic and metaphysical modal notions, and thereby grounds the Metaphysical Plenitude thesis. If this is correct, then understanding scenarios in terms of centered worlds yields epistemic intensions that satisfy the Core Thesis.

It is nevertheless useful to have an approach to the space of epistemic possibilities that is neutral on these substantive questions about metaphysical possibility. This allows even those philosophers who deny Metaphysical Plenitude to make use of the notion of an epistemic intension, and allows a maximally general defense of the epistemic understanding of two-dimensional semantics.

3.4.2 Scenarios as maximal hypotheses

The alternative is to understand scenarios in purely epistemic terms from the start. One might reasonably hold that since we want epistemic intensions to be constitutively connected to the epistemic realm, we need not invoke the metaphysical modality at all. Instead, we can do things wholly in terms of the epistemic modality. There are a couple of ways one might proceed here. One could introduce the notion of a scenario (a maximal epistemic possibility) as a modal primitive, in the same way that some philosophers introduce the notion of a world (a maximal metaphysical possibility) as a modal primitive. Or one could try to *construct* scenarios directly out of materials that are already at hand.

I take the second course in Chalmers (forthcoming), examining a detailed construction. I do not have space to do that here, but I can give a brief idea of how one might proceed. The idea I will outline is a linguistic construction of scenarios,

constructed out of linguistic expressions in an idealized language, along with a basic operator of epistemic possibility.

Let us say that a sentence D of a language L is *epistemically complete* when (i) D is epistemically possible, and (ii) there is no sentence S of L such that both D&S and D&∼S are epistemically possible. When D is epistemically complete, it is in effect as specific as an epistemically possible sentence can be. Let us say that D is *compatible* with H when D&H is epistemically possible, and D *implies* H when D&∼H is epistemically impossible (that is, when there is an a priori entailment from D to H). Then if D is epistemically incomplete, it leaves questions open: there will be H such that D is compatible with H but D does not imply H. If D is epistemically complete, D leaves no questions open: if D is compatible with H, D implies H. Note that D need not explicitly include every such hypothesis as a conjunct; these hypotheses need only be implied.

Intuitively, scenarios should correspond to epistemically complete *hypotheses*, whether or not they are expressible in a language such as English. It is likely that actual languages do not have the expressive resources to express an epistemically complete hypothesis, as they are restricted to finite sentences and have a limited lexicon. So for the purposes of this construction, we need to presuppose an idealized language that can express arbitrary hypotheses. In particular, our language L should allow infinitary sentences (at least infinitary conjunctions) and should have terms that express every possible concept, or at least every concept of a certain sort. It is also important that expressions in L are *epistemically invariant*, so that there cannot be two tokens S_1 and S_2 of the same sentence type (used with full competence) such that S_1 is epistemically necessary and S_2 is not. The exact requirements for L raise subtle issues, but we can pass over them here.

We can then focus on epistemically complete sentences of L. By the idealization, every such sentence will express a maximally specific hypothesis, and vice versa. So scenarios should correspond to epistemically complete sentences in L, although perhaps with more than one such sentence per scenario. We can say that two sentences S and T are *equivalent* when S implies T and T implies S (that is, when S&∼T and T&∼S are epistemically impossible). Any epistemically complete sentences in L will then fall into an equivalence class. We can now identify scenarios with equivalence classes of epistemically complete sentences in L. To anticipate the definition of verification: we can also say that a scenario verifies a sentence S (of an arbitrary language) when D implies S, where D is an epistemically complete sentence of L in the scenario's equivalence class.

Defined this way, scenarios are tailor-made to satisfy the Plenitude Principle. This principle requires the following:

> **Epistemic Plenitude**: For all S, if S is epistemically possible, then some epistemically complete sentence of L implies S.

Here S may be a sentence token in any language (not necessarily in L). To see the plausibility of this thesis, first note that because L has unlimited expressive power, some epistemically possible sentence S_1 of L will imply S. Second, it is plausible that any epistemically possible sentence S_1 of L is implied by some epistemically complete

sentence D of L. Intuitively, to obtain D from S_1, one simply conjoins arbitrary sentences that are epistemically compatible with S_1 (and other conjoined sentences) until one can conjoin no more. The issue is not completely trivial, as there might be endless infinitary conjunction with no maximal point, but under certain reasonable assumptions, such a sentence will exist. If so, then every epistemically possible sentence is verified by some scenario. In reverse, it is clear that any sentence verified by a scenario is epistemically possible. So the corresponding version of the Plenitude Principle is plausibly true.

In effect, this construction formalizes the intuitive idea of a maximal hypothesis: a maximal hypothesis is equivalent to an equivalence class of epistemically complete sentences in an idealized language. We might say that where the first approach takes a *metaphysical* approach to scenarios, on which they correspond to centered metaphysically possible worlds, the second approach takes an *epistemic* approach to scenarios, on which they correspond to maximal hypotheses.

What is the relationship between the two constructions? My own view is that there is a close correspondence: every centered world corresponds to a maximal hypothesis, and every maximal hypothesis corresponds to a centered world. (Not quite one-to-one: in certain cases there may be more than one centered world per maximal hypothesis, for example when there are symmetrical worlds with symmetrically corresponding centers.[9]) If so, then the Plenitude Principle will plausibly be satisfied either way. But philosophers who deny Metaphysical Plenitude will deny the close correspondence, holding that there are maximal hypotheses that correspond to no centered world. For example, a philosopher who holds that 'There is an omniscient being' is necessary but not a priori will hold that there is a maximal hypothesis that verifies the negation of the sentence in question, but that there is no centered metaphysically possible world in the vicinity. Such a philosopher should embrace the epistemic approach to scenarios.

The epistemic approach to scenarios is grounded more purely in the epistemic realm, and its central theses require fewer commitments than the metaphysical approach. For this reason, one can argue that the epistemic approach to scenarios is more basic. Centered worlds are more familiar and are useful for various applications, however, so I will use both understandings of scenarios in what follows.

On either understanding, one scenario will be privileged with respect to any statement token as the *actualized* scenario at that token. On the world-based view, this will be the world centered on the speaker and the time of utterance. On the epistemic view, this will correspond to the maximal hypothesis that is true of the world from the speaker's perspective at the time of utterance. In general, we expect that when an

[9] See Chalmers (forthcoming), section 4(4) for more on ways in which there could be more than one centered world per maximal hypothesis. Schroeter (2004) raises the possibility that there are intrinsic properties for which there is no semantically neutral conception. If there are such properties, then this is another source of a many-to-one correspondence. If such properties exist, then an epistemically complete description of a centered world may not need to specify their precise distribution. If so, then an epistemically complete description need not be ontologically complete, and more than one centered world (with different but isomorphic distributions of intrinsic properties) may correspond to the same maximal hypothesis.

expression token's epistemic intension is evaluated at the scenario that is actualized at that token, the result will be the token's extension.

3.5 Canonical descriptions

When we consider a scenario as actual, in order to evaluate an expression, we always grasp it under a description. This raises an issue. A scenario can be described in multiple ways, and it is not obvious that all such descriptions will give equivalent results. So we have to isolate a special class of *canonical descriptions* of scenarios under which they must be considered.

If we take the epistemic approach to scenarios by the second construction above, the choice will be straightforward. A scenario will correspond to an equivalence class of epistemically complete sentences. Here, we can say that a canonical description of the scenario is any sentence in the corresponding equivalence class. Because all of these sentences are equivalent under implication, they will all give the same results under verification.

If we take the metaphysical approach to scenarios, things are more complicated. Here, we require that a canonical description be a *complete neutral description* of the world. Both neutrality and completeness need explanation.

First, neutrality. To describe a world, we must choose sentences that are true of it. But will these be sentences true of the world considered as actual, or of the world considered as counterfactual? If we choose the first, there is a danger of circularity: evaluation of a world considered as actual will be defined in terms of canonical descriptions, which will be defined in terms of evaluation of worlds considered as actual. If we choose the second, there is a danger of incoherence: the framework requires that the XYZ-world verifies 'water is not H_2O', but 'Water is H_2O' is true of the XYZ-world considered as counterfactual. Either way, we need to ensure that sentences such as 'Water is H_2O' are not present within canonical descriptions of the XYZ-world.

The solution is to restrict canonical descriptions to *semantically neutral* expressions. Intuitively, a semantically neutral expression is one that behaves the same whether one considers a world as actual or as counterfactual. We cannot simply *define* a semantically neutral expression in this way, since the definition presupposes evaluation in a world considered as actual, and this evaluation (as developed here) presupposes the notion of a canonical description. But nevertheless we have a good grasp on the notion. For example, 'water' and 'Hesperus' are not semantically neutral; but 'and', 'philosopher', 'friend', 'consciousness', and 'cause' plausibly are. One could rely on our intuitive grasp of this notion for current purposes, or one could seek to define it.

One promising approach is to define such an expression as one that is not ''Twin-Earthable''. We can say that two possible individuals (at times) are twins if they are physical and phenomenal duplicates; we can say that two possible expression tokens are twins if they are produced by corresponding acts of twin speakers. Then a token is Twin-Earthable if it has a twin with a different 2-intension. This test works for many purposes. A semantically neutral term (in the intuitive sense) is never Twin-Earthable. But the reverse is not quite the case. For example, let L be an expression that functions to rigidly designate the speaker's height. Then any twin of L will have

the same 2-intension (since a twin speaker will have the same height), but L is not semantically neutral. One might respond by watering down the requirements of physical and phenomenal duplication (perhaps to some sort of mental duplication), but similar cases will still arise: e.g. if M is an expression that rigidly picks out 1 if the speaker has visual experience, and 0 if not, then M is not Twin-Earthable even by this sort of standard, but it is nevertheless not semantically neutral.[10]

A better characterization might be as follows: a semantically neutral expression is one whose extension in counterfactual worlds does not depend on how the actual world turns out (that is, on which epistemically possible scenario turns out to be actual). This is an intuitive characterization rather than a formal characterization: it invokes the intuitive idea of dependence of counterfactual extensions on the actual world, and formalizing this idea would require something equivalent to the two-dimensional framework (with ensuing circularity). But nevertheless, we have a good grip on the notion. In this sense, it is clear that most names, natural kind terms, and indexicals are not semantically neutral (and neither are L or M above), while numerous other terms (such as those listed above) are plausibly semantically neutral.

A precise formal characterization of semantic neutrality remains an open question for future research. One might try a characterization wholly in terms of our modal operators of epistemic and metaphysical necessity (that is, apriority and necessity), but it is not entirely clear how this would work. In the meantime, the intuitive characterization suffices for our purposes. It is also useful to stipulate that terms with context-dependent behavior, such as "heavy", are not semantically neutral. This allows us to describe worlds using expression types and not just expression tokens.

To characterize a centered world, semantically neutral terms must be supplemented by some indexical terms, to characterize the location of a center. The best way to do this is the following. We can say that a statement is in canonical form when it has the form D & 'I am D_1' & 'now is D_2', where D, D_1, and D_2 are all semantically neutral, and D_1 and D_2 are identifying predicates relative to the information in D (that is: D implies 'Exactly one individual is D_1' and 'Exactly one time is D_2'). We can say that a *neutral description* of a centered world is a statement in canonical form such that D is true of the world, D_1 is true of the subject at the center, and D_2 is true of the time at the center. (If the center of a centered world includes entities other than an individual and a time, then one can extend similar treatment to these entities.)

In a few cases involving completely symmetrical worlds, there may be no identifying predicates available: that is, there may be no semantically neutral predicates true only of the individual (or time) at the center. In that case, one can invoke a maximally specific predicate instead: a predicate D_1 such that for all D_2 true of the center, D entails 'everything that is D_1 is D_2'. Here, two centered worlds that differ only in

[10] Non-Twin-Earthability is related to Bealer's (1996) notion of semantic stability: "an expression is semantically stable iff, necessarily, in any language group in an epistemic situation qualitatively identical to ours, the expression would mean the same thing" (Bealer 1996, 134). It is clear that semantic stability cannot be used to characterize semantic neutrality, for the same reasons as in the case of non-Twin-Earthability. For example, the expression M in the text is semantically stable but not semantically neutral.

symmetrical placement of the center may yield the same canonical description. This is reasonable, as intuitively both worlds correspond to the same maximal hypothesis.

Second, completeness. We require that a canonical description be a *complete* neutral description of a centered world. There are two possibilities here. First, we can appeal to a criterion in terms of (metaphysical) necessity. Let us say that a semantically neutral description of a world is ontologically full when it (metaphysically) necessitates all semantically neutral truths about that world, and is minimal among the class of descriptions with this property. For example, if physicalism is true, a full semantically neutral specification of fundamental physical truths will be ontologically full. Then an ontologically complete neutral description of a centered world is a neutral description where the first (non-indexical) component of the description is ontologically full.

Alternatively, we can appeal to epistemic completeness. In this sense, a complete neutral description of a centered world is simply a neutral description that is epistemically complete. This requires the claim that for any centered world, there exists an epistemically complete neutral description. This claim is nontrivial, but there are good grounds to accept it. One can argue that although non-neutral terms are modally distinctive, they do not add fundamentally new *epistemic* power to a language, so that neutral terms constitute what I call an *epistemic basis* (see Section 3.6) for the space of epistemic possibilities.

It is not hard to see that if Metaphysical Plenitude is correct, then an ontologically complete neutral description will also be a epistemically complete neutral description.[11] If so, we can then use either criterion for a canonical description. There will arguably be more explanatory power, however, in using a complete description in the ontological sense, and then allowing this description to epistemically determine all truths about a world considered as actual.

If Metaphysical Plenitude is false, then the two criteria will not coincide. An ontologically complete neutral description will not be epistemically complete, and it will leave some hypotheses unsettled (e.g. the complete physical truth about the world may leave the Continuum Hypothesis unsettled, even if it is necessarily true). If we require that canonical descriptions be ontologically complete, the epistemic intensions of these hypotheses will have an indeterminate truth-value. A consequence may be that when an expression's epistemic intension is evaluated at the actual centered world of the expression, it does not yield the expression's extension (e.g., the epistemic intension of CH may be indeterminate at the actual world, even if CH is true). If, on the other hand, we require epistemic completeness, then the epistemic intensions of the relevant claims will have a determinate truth-value (e.g. the epistemic intension of CH will be true or false at the world according to whether CH itself is true or false there). One might do things either way, depending on

[11] The statement of Metaphysical Plenitude uses the notion of verification, which in turn requires the notion of a canonical description. For the purposes of interpreting Metaphysical Plenitude, we can assume that the canonical descriptions are required to be epistemically complete. *If* Metaphysical Plenitude formulated this way is correct, ontologically complete descriptions will give the same results as epistemically complete descriptions.

one's purposes, although for most purposes it is probably best to require epistemic completeness overall. In any case, this situation will not matter much for our purposes, since we already know that if Metaphysical Plenitude is false, then the Core Thesis will be false when scenarios are understood as centered worlds.

(A third alternative is to require "qualitative completeness", where this is characterized as in Chalmers (2002a) in terms of a notion of positive conceivability. This yields a notion that is usefully intermediate between epistemic completeness and ontological completeness. But I will leave this option aside here.)

It is clear that if scenarios are understood as centered worlds, the characterization of canonical descriptions is significantly more complicated than if scenarios are understood in wholly epistemic terms. This may be another point in favor of the purely epistemic understanding of scenarios.

3.6 Scrutability

Given the epistemic understanding of scenarios, one might have the following worry: the epistemic intension of a sentence may be well-defined, but it is *trivial*. The triviality comes from the requirement that descriptions be epistemically complete. One may worry that in order for a description to be epistemically complete, it will need to specify the truth or falsity of most sentences S explicitly. For example, 'Water is H_2O' will be true precisely in those scenarios that have 'Water is H_2O' in their canonical description, and it will be false precisely in those scenarios that have 'Water is not H_2O' in their canonical description. If this sort of thing is typical, then epistemic evaluation as defined will have an uninteresting structure.

A related worry arises on the metaphysical understanding of scenarios. Here, the issue concerns the thesis (mentioned in the previous section) that there is an epistemically complete neutral description of any centered world. If one had the worry just mentioned about 'Water is H_2O', one might worry that an epistemically complete description of a centered world requires non-neutral terms, such as 'water'. The key question is whether the truth-value of all sentences S is epistemically necessitated by a description of a centered world in terms of semantically neutral expressions plus indexicals. If this is not the case, then as defined, the epistemic intension of the relevant sentences will be indeterminate at the relevant centered worlds.

These worries are reasonable enough, but I think that they are ultimately unfounded. In what follows, I will concentrate on the worry that applies to the epistemic understanding, but similar considerations also apply to the metaphysical understanding. To answer the worry, one needs to make the case that epistemically complete descriptions do not need to specify the truth or falsity of most statements explicitly, so that epistemic evaluation does not have a trivial structure. To see this, it is useful to focus on the *actual* world, and consider what an epistemically complete description of this world must contain. The sort of argument I give here is presented in much more depth by Chalmers and Jackson (2001) and Chalmers (2002a); but here I will give the basic idea.[12]

[12] Chalmers and Jackson (2001) can be seen as providing a crucial part of the foundation for the two-dimensional framework as it is understood here, even though the framework is hardly

The second principle underlying the epistemic understanding of the two-dimensional framework was what we might call the *scrutability of truth*. This can be put informally as the thesis that once we know enough about the state of the world, we are in a position to know the truth-values of our sentences. Furthermore, we usually need not be informed about a sentence explicitly in order to know whether it is true. We could put this somewhat more precisely as follows:

> **Scrutability of Truth**: For most terms T used by a speaker, then for any truth S involving T, there exists a truth D such that D is independent of T, and such that knowing that D is the case puts the speaker in a position to know (without further empirical information, on idealized rational reflection) that S is the case.

Here, we can say that D is independent of T when D does not contain T or any close cognates. Of course this notion is somewhat vague, as is the notion of "most" above, but this does not matter for our purposes. To save breath, we can abbreviate "knowing that D is the case puts the speaker in a position to know (without further empirical information, on idealized rational reflection) that S is the case" as "D is epistemically sufficient for S".

Take the case of 'water'. Here, we can let D be a truth specifying an appropriate amount of information about the appearance, behavior, composition, and distribution of objects and substances in one's environment, as well as information about their relationship to oneself. D need not contain the term 'water' at any point: appearance can be specified in phenomenal terms, behavior and distribution in spatiotemporal terms, composition in microphysical or chemical terms. Then D is epistemically sufficient for 'Water is H_2O'. When one knows that D is the case, one will be in a position to know all about the chemical makeup of various liquids with various superficial properties in one's environment, and will thereby be able to infer that water is H_2O. After all, this information about appearance, behavior, composition, and distribution is roughly what we need in the case of ordinary knowledge, to determine that water is H_2O. And there is no need for further empirical information to play a role here: even if we suspend all other empirical beliefs, we can know that *if* D is the case, then water is H_2O.

The same goes for terms such as 'Hesperus'. Once again, if D contains appropriate information about the appearance, behavior, composition, and distribution of various objects in the world, then D is epistemically sufficient for 'Hesperus is Venus', for 'All renates are cordates', and so on. The information in D enables one to know that the object that presents a certain appearance in the evening is the same as the object that presents a certain appearance in the morning, and so enables us to know that Hesperus is Phosphorus. Something similar applies to 'heat', 'renate', and so on.

mentioned in the paper (which is packaged as a response to Block and Stalnaker on the explanatory gap). Section 3 of the paper in effect argues for the scrutability thesis in a general form, and sections 4 and 5 defend a specific version of the thesis. The reply to objection 6 in section 5 is particularly important in defending the a priori entailment version of the scrutability thesis. Sections 8 and 9 of Chalmers (2002a) provide a further defense of a version of the thesis.

Here, the base information need not contain terms such as 'Hesperus' or 'renate', or any cognates. And no further empirical information is required: the information in the base is all that is needed.

Something similar applies for terms like 'philosopher', or even names like 'Gödel' or 'Feynman'. Here, the base information D may need more than in the cases above: for example, it may need to include information about people and their mental states, and the use of certain names, and so on. For example, once I know enough about the history of the use of the name 'Gödel' by others in my community, about the properties of relevant individuals, and so on, then I will be in a position to know that Gödel was a mathematician, even if I had no substantive knowledge of Gödel beforehand. And again, my information need not use the terms 'Gödel' or 'mathematician' to do this. It might use the quite different term '' 'Gödel' '', in order for me to track down the referent via those from who I obtained the name, but that is legitimate in this context.

This pattern may not apply to *all* expressions. There are plausibly some primitive terms (perhaps 'and', 'cause', and 'conscious', for example) such that to know whether a sentence involving these terms is true, one needs a base that includes those terms or relevant cognates that invoke them implicitly. But as long as the principle applies reasonably widely, it is good enough.

By the sort of reasoning above, one can infer a slightly stronger claim. Let us say that a vocabulary is a set of terms, and that a V-truth is a truth that uses only terms in V. Then we can say: there is a relatively limited vocabulary V such that for any truth S, there is a V-truth D such that D is epistemically sufficient for S. To arrive at V, intuitively, we might simply eliminate terms one by one from the language according to the scrutability principle laid out above, until we cannot eliminate any further. Exactly how limited V must be is an open question, but I think the sort of reasoning above gives good ground to accept that it will involve only a small fraction of the original language. One can put the claim in a slightly stronger form:

> **Scrutability of Truth II**: There is a relatively limited vocabulary V such that for any truth S, there is a V-truth D such that D implies S.

Here, we have moved from ''D is epistemically sufficient for S'' to ''D implies S'': that is, that the material conditional 'D $\supset$ S' is a priori. This is a stronger but not a vastly stronger claim, given that epistemic sufficiency involved ''no further empirical information''. One can argue for it along much the same lines as above, suggesting that even a speaker who suspends all empirical beliefs can know that *if* D is the case, then S is the case. Chalmers and Jackson (2001) argue in much more depth that this sort of conditional is a priori (for a specific choice of V). A point made there is worth noting here: this sort of a priori entailment does not require that there is an explicit definition of the terms in S using the terms in V.[13]

[13] For this reason, the epistemic two-dimensional framework set out here does not require or entail that the epistemic intension of an expression be analyzable in terms of some explicit description: for example, it is not required that a name or a natural kind term N be analyzable

(Note that even if one is skeptical about apriority, the general point about epistemic sufficiency is still plausible. Such a skeptic can instead appeal to an alternative notion of epistemic necessitation, such as one understood in terms of rational inference. Corresponding theses about scrutability and nontriviality will remain plausible given such a notion.)

It is also plausible that there is some V-truth D that implies all V-truths. Of course D may need to be an infinitary conjunction, but we may as well stipulate that V is part of our idealized language, so this is no problem. We can think of D as a conjunction of the simple V-truths about the world, or as a conjunction of all V-truths of up to a certain level of complexity. There is plausibly a level such that any more complex V-truth will be implied by this sort of conjunction. If so, it follows that D implies all truths about the world. It follows plausibly that D is epistemically complete (if D is compatible with H and ∼H, then all truths about the world are compatible with H and with ∼H, which is plausibly impossible).

Exactly what is required for the vocabulary V and the description D is an open question. Chalmers and Jackson (2001) and Chalmers (2002a) argue that a specific description D will work here: PQTI, the conjunction of microphysical and phenomenal truths with certain indexical truths and a "that's-all" truth. If this is right, then V requires only the vocabulary required for PQTI. It is possible that the vocabulary might be stripped down further, if Q is implied by P (as some physicalists hold), or if P is implied by a description in a more limited vocabulary, such as one in terms of space, time, and causal connections (an appropriate Ramsey sentence, for example). But in any case, this specific claim is not required here. The only claim required is that *some* limited vocabulary V suffices for this purpose.

What goes for the actual world goes also for any epistemic possibility. There is nothing special about the actual world here. Given any class of epistemically compatible sentences in our idealized language, one can strip down the vocabulary involved in it in the same sort of way as before, until one has a limited vocabulary V′ such that each of the original sentences is implied by a V′-sentence. It follows by similar reasoning to the above that for any scenario W, there will be a limited vocabulary V′ such that there is an epistemically complete V-truth that corresponds to the scenario. Of course the vocabulary may differ between scenarios. For example, there are presumably epistemically possible scenarios that involve conceptually basic kinds that are alien to our worlds. If so, the vocabulary required to describe our world must be expanded to describe this scenario. But the resulting vocabulary will still be limited.

as 'the actual D' for some description D. Likewise, an expression's epistemic intension need not correspond directly to any descriptive belief of the speaker: for example, it is not required that one who uses a term N has a priori "identifying knowledge" to the effect the referent of N is φ, for some property φ. All that is required is that certain conditionals be epistemically necessary.

This bears on criticisms of two-dimensionalism raised by Soames (2004) and by Byrne and Pryor in this volume. A number of Soames's arguments rest on criticizing the thesis that names are analyzable as rigidified descriptions. The central arguments of Byrne and Pryor rest on criticizing the thesis that users of names and natural kind terms have a priori identifying knowledge. The framework I have outlined is not committed to these theses (in fact, I think that the theses are probably false), so the corresponding arguments do nothing to undermine the framework I have outlined.

Let us say that a *basic* vocabulary is a minimal vocabulary V′ such that every epistemically possible sentence is implied by some V′-sentence. We can think of such a vocabulary as providing an *epistemic basis*: the terms in it express a set of concepts sufficient to cover all of epistemic space. Given the reasoning above, there is reason to believe that a basic vocabulary will be a relatively limited vocabulary. Exactly how small a basic vocabulary can be is again an open question, but it may well involve only a very small fraction of the terms of the original language. With such a vocabulary in place, we can think of a scenario as corresponding to an equivalence class of epistemically complete V′-sentences, rather than of arbitrary epistemically complete sentences.

(Note that there is no need to appeal to a basic vocabulary for the *definition* of epistemic intensions. The canonical descriptions invoked in the definition are not restricted to a basic vocabulary, although it is easy to see that any such description will be epistemically equivalent to a description in a basic vocabulary.)

If a reasonably limited basic vocabulary exists, it follows that epistemic intensions are nontrivial. An epistemically complete description need not specify the status of most sentences explicitly. Most terms, such as 'water' and 'H_2O', will plausibly not be required in a basic vocabulary, so sentences involving these terms will be *nontrivially* true or false in scenarios. For all we have said here, it may be that some claims (for example 'there is space') are in a sense trivially true in some scenarios and trivially false in others, but this is only to be expected: it is analogous to the trivial truth or falsity (in an analogous sense) of claims about ontologically fundamental properties in metaphysically possible worlds. So there will be plenty of interesting structure to epistemic intensions in general.

3.7 Subsentential epistemic intensions

So far I have defined epistemic intensions only for sentences. It is not too hard to define them for subsentential expressions, such as singular and general terms, kind terms, and predicates, but there are a few complexities. I will take it that we have already decided on independent grounds what sort of extensions these expressions should have: e.g. individuals, classes, kinds, and properties. Differences choices could be made here, but the same sort of treatment will work.

The details depend to some extent on whether we take the metaphysical or the epistemic approach to scenarios. The difference is that centered worlds already come populated with individuals and the like (or at least we are familiar with how to regard them as so populated), whereas maximal hypotheses do not (or at least we are less familiar with how to populate them).

If we take the metaphysical approach to scenarios: let W be a centered world with canonical description D, and let T be a singular term. In most cases, D will imply a claim of the form 'T = T*', where T* is a semantically neutral singular term. (Here I include definite descriptions as singular terms.) If so, the epistemic intension of T picks out the referent of T* in W (that is, it picks out the individual that T* picks out when W is considered as counterfactual). In some symmetrical worlds, it may be that there is no such semantically neutral T*, but there is a T* that involves semantically neutral terms plus 'I' and 'now' (plus other basic indexicals, if any). In this case, one

can replace the indexicals in T* by labels for the entities at the center of the world, yielding an expression T** such that the epistemic intension of T picks out the referent of T** in W. If there is no such T*, then the epistemic intension of T is null in W.

One can do the same for general terms, appealing to claims of the form 'For all x, x is a T iff x is a T*', and holding that the epistemic intension of T in W picks out the referent of T*, for a T* that is semantically neutral (perhaps plus indexicals). For kind terms, we again appeal to identities 'T is T*'. For predicates, we can appeal to claims of the form 'For all x, x is T iff x is T*'. This delivers extensions for the epistemic intensions straightforwardly.

If we take the epistemic view of scenarios, then we need to populate scenarios with individuals and the like. If we simply admit scenarios as a basic sort of abstract object with certain properties, one could simply stipulate that they contain individuals that can serve as the extensions of relevant expressions—much as many of those who introduce possible worlds simply stipulate something similar. But it is useful to go through an explicit construction.

Let W be a scenario with canonical description D. Let us say that two singular terms T_1 and T_2 are equivalent under W if D implies 'T_1 is T_2'. Then we can identify every equivalence class of singular terms under W with an individual in W, and hold that the epistemic intension of T in W picks out the individual corresponding to T's equivalence class in W. As for general terms: every general term G will pick out a class of individuals. One of the individuals defined above will be in G precisely when D implies 'T is a G', for some T that picks out the individual. One can do something similar for predicates and kind terms: the details will depend on the precise view one takes of properties and kinds and their relation to individuals, so I will not go into them here.

There is one worry: what if the truth of certain existentially quantified claims in a scenario requires individuals that are not the referent of any singular term? For example, there may be a predicate φ such that D implies '$\exists x\, \varphi(x)$', and D does not imply any claim of the form '$\varphi(T)$', where T is a singular term. Of course since D is epistemically complete, it will tell us exactly how many individuals have φ, whether some individuals with φ also have ψ and some do not, and so on. It is not hard to see that this sort of case will ultimately require predicates φ (perhaps an infinitely conjunctive predicate) such that D implies that there exists more than one individual with ψ, and such that for all predicates ψ, D implies that these individuals are indistinguishable with respect to ψ. In this case, the individuals will be indistinguishable even in our idealized language, presumably because of deep symmetries in the world. In such a case, if D implies that there are n individuals with φ, one can arbitrarily construct n individuals, perhaps as ordered pairs $(\varphi', 1) \ldots (\varphi', n)$, where φ' is the equivalence class containing φ, and stipulate that all of these individuals fall under the extension of φ, and of other predicates and general terms as specified by the relevant D-implied universally quantified truths about individuals with φ.

In this way, we can construct the relevant classes of individuals and the like, and specify the extensions of various expressions' epistemic intensions. The construction ensures that where the extension of a complex expression is a compositional function

of the extensions of its parts, then the same will be true of the extension of a complex expression relative to a scenario. For an identity (e.g. '$T_1 = T_2$'), compositionality will be ensured by the equivalence class construction. For a predication (e.g. 'T is a G', or Ï† (T)), this will be ensured by the appropriate construction of extensions for general terms (as above) or predicates. The machinations two paragraphs above ensure that existential quantification will work straightforwardly, and universal quantification is guaranteed to work (if D implies $\forall x\ \varphi(x)$, then every individual constructed above will have φ). Logical compositionality is guaranteed at the sentential level (if D implies both S and T, D will imply S&T, and so on). So the epistemic intension of a complex expression will be a compositional function of the epistemic intension of its parts.

Of course once one has engaged in this sort of construction, one need not usually bother with the details again. It is perfectly reasonable thereafter to speak of a scenario as containing individuals and the like, and to speak about terms as picking out various individuals in a scenario, quite independently of the details of the construction. On the epistemic approach to scenarios, for most purposes one can think of them as abstract objects that may behave somewhat differently from possible worlds, but that have the same sort of status in our ontology.

3.8 Tokens and types

As I have approached things, epistemic intensions have been assigned to expression tokens rather than expression types (such as linguistic types). The reason for this is straightforward. It is often the case that two tokens of the same linguistic type can have *different* epistemic intensions. This difference arises from the fact that different speakers may use the same expression so that it applies to epistemic possibilities in different ways. And this difference arises in turn from the fact that different speakers may use the same term with different a priori connections.

For example, it is often the case that two speakers will use the same *name* with different a priori connections. The canonical case is that of Leverrier's use of 'Neptune', which he introduced as a name for (roughly) whatever perturbed the orbit of Uranus. For Leverrier, 'If Neptune exists, it perturbs the orbit of Uranus' was a priori. On the other hand, later speakers used the term (and still do) so that this sentence is not a priori for them: it is epistemically possible for me that Neptune does not perturb the orbit of Uranus. We can even imagine that when Leverrier's wife acquired the name, she did not acquire the association with Uranus, so that she is in no position to know the truth of this sentence a priori.

How can we characterize the epistemic intension of Leverrier's tokens of 'Neptune'? To a first approximation, we can say that in any scenario, Neptune picks out whatever perturbs the orbit of Uranus in that scenario. How can we characterize the epistemic intension for Leverrier's wife? This is a bit trickier, but we can assume that for his wife to determine the reference of 'Neptune', she would examine Leverrier's own use and see what satisfies it. So to a first approximation, his wife's epistemic intension picks out whatever Leverrier refers to as 'Neptune' in a given scenario. One can find a similar (although less stark) variation in the epistemic intensions of many names, and perhaps natural kind terms.

Something similar applies to many uses of context-dependent terms, such as 'heavy'. What I count as heavy varies with different uses of the term. In some contexts, 'My computer is heavy' may be true, and in other contexts it may be false, even though it is the same computer with the same weight. Correspondingly, the way I apply a term across epistemic possibilities will vary with these uses: if I *suppose* that my computer weighs such-and-such, I may hold the utterance true in the first case but not the second.

As we have defined epistemic intensions, they are grounded in the behavior of sentences under an epistemic necessity operator. So the variation in epistemic intensions of two expressions of the same type is traceable to variations in the epistemic necessity of two type-identical sentences. In particular, it will be traceable to variations in the apriority of two type-identical sentences. And this variation is traceable to variations in the apriority of the thoughts that the two sentences express.

Here we need to say a little more about thoughts. A thought is understood here as a token mental state, and in particular as a sort of occurrent propositional attitude: roughly, an entertaining of a content. The idea is that this is the sort of propositional attitude that is generally expressed by utterances of assertive sentences. Such utterances typically express occurrent beliefs, but they do not always express occurrent beliefs, as subjects do not always believe what they say. Even in these cases, however, the subject *entertains* the relevant content: a thought is an entertaining of this sort. Like beliefs, thoughts are assessable for truth. Thoughts can come to be *accepted*, yielding beliefs, and thoughts can come to be *justified*, often yielding knowledge. When an utterance expresses a thought, the truth-values of the utterance and the thought always coincide.

On this way of approaching things, we assume a relation of expression between statements and thoughts, and we assume a notion of epistemic necessity as applied to thoughts. The latter notion might be seen as the true conceptual primitive of the approach. On the account where epistemic necessity is tied to apriority, we can characterize it further by saying: a thought is epistemically necessary when it can be justified independently of experience, yielding a priori knowledge. We can then say that a thought is epistemically possible when its negation is not epistemically necessary. Two thoughts are epistemically *compatible* when their conjunction is epistemically possible. One thought *implies* another when the first is epistemically incompatible with the negation of the second. Here we assume that two thoughts of the same subject can stand in a relation of negation, and that a thought can stand in a relation of conjunction or disjunction to a set of two or more other thoughts of the same subject.

With these notions in hand, we can characterize epistemic necessity and necessitation as applied to sentences. A sentence token is epistemically necessary iff it expresses an epistemically necessary thought. A sentence type is epistemically necessary iff any token of the type (used competently and literally) is epistemically necessary. If D and E are sentence types, we can say that D epistemically necessitates E when D&$\sim$E is epistemically impossible. If D is a sentence type and S is a sentence token: let us say that a thought is a D-thought if it is the sort apt to be expressed by D. Then D epistemically necessitates S when a possible D-thought of the subject will imply the thought expressed by S. Equivalently, D epistemically necessitates S when, if the

thought expressed by S were to be disjoined with a ~D-thought, the resulting thought would be epistemically necessary.

We can also use this framework to directly define epistemic intensions for thoughts as well as utterances. Much as above, we can say that a scenario verifies a thought when disjunction of the thought with a ~D-thought is epistemically necessary, where D is a canonical description of the scenario. This yields a notion of mental content that can be applied to beliefs, thoughts, and any propositional attitude with a mind-to-world direction of fit (see Chalmers 2002c). Using the definitions above, we can see that when an utterance expresses a thought, the epistemic intension of the utterance will be identical to the epistemic intension of the thought.

This framework enables us to see how two tokens of the same type can differ in apriority. When Leverrier says 'If Neptune exists, it perturbs the orbit of Uranus', his statement presumably expresses a priori knowledge, and certainly expresses a thought that can be justified a priori. If his wife utters the same sentence, no amount of a priori rational reflection alone could justify the thought she expresses. Similarly, it is possible that two names for a single individual ('Bill Smith' and 'William Smith') are used completely interchangeably by one person, so that an utterance of an identity statement involving the names expresses a trivial thought. Such an utterance will then be a priori. But there clearly may be others for whom a corresponding utterance expresses a nontrivial thought and is a posteriori.[14] Finally, if I say 'someone with 1,000 hairs on their head is bald' on one occasion, it may express an a priori false thought (one whose negation is a priori justifiable), while if I say it on another occasion, it may express a thought that is not a priori false, and may be plausibly true.

In a similar way, this framework enables us to see how two tokens of the same type can have different epistemic intensions. Let S be the sentence 'Neptune is an asteroid', and let D be a canonical description of a scenario W in which the orbit of Uranus is perturbed by an asteroid and in which no-one has ever used the term 'Neptune'. (We can abstract away from complications involving the intension of 'Uranus' and 'asteroid'.) Then D epistemically necessitates Leverrier's utterance of S: a thought that D obtains would imply the thought Leverrier expresses with S. But D does not epistemically necessitate Leverrier's wife's utterance of S: a thought that D obtains would not imply the thought that his wife expresses with S. (Note that D itself will not exhibit this sort of variation, as expressions in the idealized language are required to be epistemically invariant.) So Leverrier's utterance of S is verified by W, while his wife's utterance is not. So the two utterances have different epistemic intensions.

One might reasonably ask: in languages such as English, what sorts of simple terms have epistemic intensions that vary between speakers and occasions of use? This happens most clearly for: (i) proper names (such as 'Neptune' and 'Gödel'); (ii) ordinary natural kind terms (such as 'water' and 'gold'); (iii) demonstratives (such as 'that' and 'there'); and (iv) many context-dependent terms (such as 'heavy' and 'bald'). For terms like this, it is clear that an epistemic intension is not part of a term's "standing meaning", where this is understood as the sort of meaning that

[14] This sort of case is discussed in Chalmers (2002b), section 9.

is common to all tokens of a type in a language. Instead, it is a sort of "utterance meaning" or "utterance content". Some theorists use the term "meaning" only for standing meaning, but this is a terminological matter. The substantive point is that the framework yields a useful and interesting sort of semantic value in the broad sense, one that can be associated with utterances and that can play a useful explanatory role. (There is more discussion of this matter in Chalmers 2002b, section 8.)

Are there terms for which an epistemic intension is common to all tokens of a type? This is perhaps most plausible for certain indexicals, such as 'I' and 'today' (at least setting aside unusual uses, and any context-dependence at the boundaries). It may also hold for some descriptive terms, such as 'circle'. Most of these have some context-dependence, but this can be regimented out more straightforwardly than the epistemic variability of names and natural kind terms. Finally, it may hold for some descriptive names (e.g. 'Jack the Ripper'), at least for a certain period of their existence. For terms like these, an epistemic intension might be seen as part of their standing meaning.

3.9 Apriority

On the main approach advocated here, epistemic necessity is regarded as a sort of apriority. This requires us to say a bit more about the notion of apriority. There are various ways in which apriority can be understood, but current purposes require a fairly specific understanding of it. A characterization of the relevant notion of apriority might run something like this. A sentence token is a priori when it expresses an a priori thought. A thought is a priori when it can be conclusively non-experientially justified on ideal rational reflection.

There are five distinctive features of this conception of apriority that deserve comment. The first feature is that the relevant sort of apriority is *token-relative*. The second is that apriority is *mode-of-presentation-sensitive*. The third is that apriority is *idealized*. The fourth is that apriority is *non-introspective*. The fifth is that apriority is *conclusive*. The first feature was been discussed in the previous section. Here I will say a little about the other four.

Mode-of-presentation sensitivity: Intuitively, sentences such as 'Hesperus is Phosphorus' are a posteriori. But some theorists (e.g. Salmon 1986; Soames 2002) hold that such sentences are a priori, on the grounds that they express trivial singular propositions that can be known a priori (e.g. by knowing that Venus is Venus). On the current definition of apriority, tokens of such a sentence are not a priori. The thought expressed by an utterance of 'Hesperus is Phosphorus' clearly cannot be justified independently of experience. At best, a different thought associated with the same singular proposition can be so justified. So on the current definition, the utterance is not a priori. On this approach, as on the intuitive understanding, apriority is sensitive to mode of presentation. The apriority of an utterance is grounded in the epistemic properties of a corresponding thought, which are tied to the inferential role of that thought in cognition.[15]

[15] Note that to say that a token of a sentence S produced by speaker A is a priori is *not* to say that a knowledge ascription of the form 'A knows a priori that S' (or 'A can know a priori that

Idealization: Here the notion of apriority is understood so that it idealizes away from a speaker's contingent cognitive limitations. A sentence token (of a complex mathematical sentence, for example) may be a priori even if the speaker's actual cognitive capacities are too limited to justify the corresponding thought a priori. What matters is that the thought could be justified a priori on idealized rational reflection.[16]

Non-introspectiveness: On some conceptions of apriority, introspective knowledge (for example my knowledge that I am thinking, or my knowledge that I believe I am Australian) qualifies as a priori. On the current conception, introspective knowledge does not qualify as a priori. We can stipulate that experiential justification should be understood in such a way to include both perceptual and introspective justification. It follows that in excluding experiential justification, apriority rules out both perceptual and introspective justification.

Conclusiveness: On the current conception, a priori justification must meet the sort of conclusive standard associated with proof and analysis, rather than the weaker standard associated with induction and abduction. On this conception, an inductive generalization from instances each of which is known a priori does not possess the relevant sort of a priori justification (even though it might be held to be a priori in some reasonable sense). Likewise, neither does an abductive conditional from total evidence to a conclusion that is grounded in and goes beyond the evidence. Intuitively, in such cases one may have non-experiential justification for believing a conclusion, but one is unable to conclusively rule out the possibility that the conclusion is false. As understood here, apriority is tied to the sort of justification that conclusively rules out the possibility that the relevant sentence is false.[17]

Of the five features above, the first two are necessary in order to capture the close tie between apriority and rational significance: it seems clear that rational significance is token-relative and mode-of-presentation sensitive. The last three are necessary in order to capture the idea that apriority should correspond to a sort of epistemic

S') is true. (Clearly a token of 'If I exist and am located, I am here' may be a priori for a speaker even if that speaker cannot know a priori that if I exist and am located, I am here. The criteria may also come apart in cases where ascriber and ascribee associate different modes of presentation with the expressions in S.) The current construal of apriority requires no commitment on the semantics of attitude ascriptions: what I have said here about the non-apriority of 'Hesperus is Phosphorus' is even consistent with Salmon and Soames' counterintuitive Millian semantics for attitude ascriptions, on which 'A knows a priori that Hesperus is Phosphorus' is true.

[16] This characterization of the idealized a priori should be seen as an intuitive characterization of a notion that is being taken as primitive. It is best not to define idealized apriority in terms of possible justification, both because of the proliferation of primitive notions, and because it could lead to problems on views on which certain conceivable cognitive capacities are not metaphysically possible. For example, if it turns out that there are strong necessities entailing that no being can construct a proof with more than a million steps, then a statement whose proof requires more steps than this will not satisfy the putative definition, but it will still count as epistemically necessary in the idealized sense I am invoking here.

[17] Again, this intuitive characterization should not be understood as an analysis of conclusive justification. It merely points to an intuitive distinction. A more detailed characterization might analyze conclusive justification of a belief in terms of the nonexistence of certain sorts of skeptical hypotheses under which the belief would be false.

necessity. We want epistemic necessity to capture the intuitive idea that some thoughts are true under all coherent hypotheses about the actual world. Inductive knowledge and introspective knowledge do not have this property (intuitively, they are false in some scenarios), while idealized mathematical knowledge does have this property. So our conception of apriority should exclude the first two and include the third.

This conception of apriority should be understood as stipulative. One can define 'a priori' in different ways, so that it is type-relative, or so that it is not sensitive to modes of presentation, or that it is not idealized, or so that it allows introspective knowledge or inductive knowledge. There is no need to adjudicate the terminological question of which of these conceptions is the "correct" one. In fact, nothing rests on the use of the term "a priori": one could simply use the term "epistemically necessary" for the stipulated notion throughout.

3.10 The second dimension

I have concentrated almost wholly on the first dimension of the two-dimensional framework. This is because the second dimension is already well-understood. But I will say a few words about it here. It is worth examining how it can be understood in a way that is parallel to the way we have understood the first dimension.

Like the first dimension, the second dimension is founded on a certain sort of possibility and necessity. For the first dimension, this is epistemic possibility and necessity, tied to what *might be* the case. For the second dimension, this is what we might call *subjunctive* possibility and necessity, tied to what *might have been* the case.

We can say that S is subjunctively possible when it might have been the case that S (more strictly, when an utterance of 'it might have been the case that S' by the speaker, with the modal operator adjusted for the relevant language, would be true). Kripke is explicit that this is the basic notion of possibility and necessity with which he is working, and almost all of his modal arguments are directly grounded in intuitions about what might have been the case.

With this basic modal operator in hand, we can proceed as before. For example, one can define a *subjunctively complete* sentence parallel to the way we defined an epistemically complete sentence. One can construct equivalence classes of subjunctively complete sentences in an idealized language. One can identify these classes as *maximal metaphysical possibilities*, or as *possible worlds*. One can give possible worlds *canonical descriptions*, which will be subjunctively complete sentences in their equivalence class.

Just as one can consider a scenario as actual, by supposing that it actually obtains, one can consider a world as counterfactual, by supposing that it had obtained. That is, instead of thinking "if D is the case, then . . .", one thinks "if D had been the case, then . . ." (where D is a canonical description of a world W). For example, for a given sentence S, one can entertain and evaluate the subjunctive conditional: "if D had been the case, would S have been the case?". In some cases, the answer will intuitively be yes: in this case, we can say that W *satisfies* S. This is a distinctive sort of *counterfactual* evaluation. When W satisfies S, we can say the *subjunctive intension* of S is true at W.

For example, if we accept Kripke's intuitions, then we will say: if the bright object visible in the evening had been Mars, then it would not have been the case that Hesperus was Mars (Hesperus would still have been Venus). In this way, our subjunctive intuitions are quite different from our epistemic intuitions. Likewise, if we accept Putnam's intuitions, then we will say: if the clear liquid in the oceans and lakes had been XYZ, then it would not have been the case that water was XYZ (water would still have been H_2O). If we accept these intuitions, we will say that the subjunctive intension of 'Hesperus is Venus' is true at all worlds (or at all worlds where Venus exists), as is the subjunctive intension of 'Water is H_2O'. These intensions differ markedly from the epistemic intensions of 'Hesperus is Venus' and 'Water is H_2O', both of which are plausibly false at many scenarios.

The subjunctive intension of a sentence S is a function from worlds to truth-values, true at W if and only if W satisfies S. Satisfaction can be intuitively characterized as above. Formally, we can say that W satisfies S when D subjunctively necessitates S, where D is a canonical description of W. We could define subjunctive necessitation by the subjunctive conditional heuristic above. Or more formally, one might say that D subjunctively necessitates S when D&$\sim$S is subjunctively impossible.

With a possible world as constructed, we can construct a space of individuals much as we did with scenarios. We can then define subjunctive intensions for subsentential expressions straightforwardly. Subjunctive intensions are defined in the first instance for expression tokens, since subjunctive necessity judgments can vary between tokens of a type. For some expression types, all tokens of the type will have the same subjunctive intension: this is arguably so for names and natural kind terms (for example 'Hesperus' and 'water'), logical and mathematical terms, and some descriptive terms (for example 'circle'). For other expression types, subjunctive intensions will vary between tokens of the type: this is so for indexicals (for example 'I') and many context-dependent terms (for example 'heavy'). In the first case, subjunctive intension may be an aspect of linguistic meaning; in the second case, it is not.

The basic ideas here are parallel between the two cases. The explicit construction of possible worlds and the like may seem like unnecessarily heavy weather; but this seems so only because possible worlds are more familiar. Perhaps one does not really need any such construction to legitimize the appeal to possible worlds; but if so, the same applies to scenarios. In both cases, one takes a modal notion as basic, and invokes a corresponding modal space as a tool of analysis.

There is one important difference between worlds and scenarios. We have a means of reidentifying individuals across worlds, but in general there is no such means of reidentifying individuals across scenarios. In the case of worlds, these claims are grounded in *de re* subjunctive intuitions of the form 'x might have been F'—read so that they are distinct in their form from de dicto subjunctive intuitions such as 'it might have been that T was F' where T denotes x. We can use these claims in conjunction with the construction above to identify certain objects in alternative possible worlds as identical to certain objects in the actual world (or alternatively, to identify objects with equivalence classes across worlds; or at least to set up counterpart relations across worlds). There is no clear analog of a *de re* modal intuition in the epistemic case: 'Hesperus is the evening star' may be a priori, but it is not clear

what it means to say that Hesperus (i.e. Venus) is such that it is a priori that it is the evening star.

In the subjunctive case, one can also ground the reidentification of individuals across worlds in de dicto subjunctive intuitions involving a privileged class of designators, that is, names. Judgments of the form 'it might have been that N was F' where N is a name for the relevant object, arguably give the same result for any name of the object, and if so can ground a sort of crossworld identification. In the epistemic case, there is in general no analog to this privileged class of designators: different names for an individual are not generally a priori equivalent, so come apart in different scenarios, and there is no way in general to isolate a privileged class of epistemically equivalent designators here. At best this may be possible in special cases, such as canonical designators for phenomenal states and abstract entities. A consequence is that quantified modal claims will not generally be well-defined in the epistemic case, and quantified modal logic will be largely inapplicable in this domain.

In many cases, a term's subjunctive intension will depend on its actual extension, or on other aspects of the actual world. This is particularly clear in the case of rigid designators such as names and indexicals. If Kripke is correct, these pick out the same individual in all possible worlds, and so pick out the term's actual extension in all possible worlds (for example, the subjunctive intension of 'Hesperus' will pick out Venus in all worlds). In these cases, the subjunctive intension of a term itself depends on the character of the actual world. Here, in effect, a term's subjunctive intension depends on which epistemic possibility turns out to be actual.

One can naturally encapsulate this behavior in a *two-dimensional intension*. This can be seen as a mapping from scenarios to subjunctive intensions, or equivalently as a mapping from (scenario, world) pairs to extensions. We can say: the two-dimensional intension of a statement S is true at (V, W) if V verifies the claim that W satisfies S. If D_1 and D_2 are canonical descriptions of V and W, we say that the two-dimensional intension is true at (V, W) if D_1 epistemically necessitates that D_2 subjunctively necessitates S. A good heuristic here is to ask "If D_1 is the case, then if D_2 had been the case, would S have been the case?". Formally, we can say that the two-dimensional intension is true at (V, W) iff '$\Box_1(D_1 \supset \Box_2(D_2 \supset S))$' is true, where '$\Box_1$' and '$\Box_2$' express epistemic and subjunctive necessity respectively. One can define two-dimensional intensions for subsentential expressions by an extension of this idea.

One complication: the construction so far makes the space of possible world derive from subjunctive modal claims. The truth of some subjunctive modal claims depends on the character of the actual world, so one might think that the space of possible worlds will do so as well (that is, the space of W's may depend on V). Whether this will be so depends on further substantive philosophical issues.

If (i) every possible world can be completely specified by semantically neutral terms (and if this is a priori), then one can require that canonical descriptions be given in these terms, and can use these descriptions to identify worlds across the spaces corresponding to each scenario. If also (ii) the truth of subjunctive modal claims in semantically neutral language is a priori, so that it does not depend on which scenario is actual, then one can identify the spaces themselves. If (i) holds but not (ii), then while we can identify worlds across spaces, some of these spaces will differ in their

extent. (Note that if Metaphysical Plenitude is a priori, as I hold, then this option is excluded.) If (i) does not hold (for example because there are pure haecceitistic differences between worlds, or because there are fundamental intrinsic properties that cannot be specified in semantically neutral terms), then it is probably best to see each scenario as being associated with a relativized space of possible worlds (putative worlds, in the case of non-actual scenarios). In this case, canonical descriptions of worlds on the second dimension will sometimes use non-neutral language, and the worlds will not always be identifiable across spaces.

In effect, this two-dimensional structure will represent the space of epistemic possibilities concerning the space of metaphysical possibilities. If (i) and (ii) hold, the extent and nature of the space of metaphysical possibilities will be determined a priori, so that we will have the same space of worlds corresponding to every scenario. If (i) or (ii) fail to hold, then the space of metaphysical possibilities will depend to some extent on which epistemic possibility turns out to be actual, and we may have different spaces of (putative) worlds corresponding to different scenarios.

Note that this worry does not affect the earlier use of semantically neutral descriptions of centered worlds for epistemic purposes. The cases where semantically neutral resources do not fully describe a world will generally correspond to cases where two centered worlds have the same canonical description for epistemic purposes, so that they correspond to a single maximal hypothesis (and to a single scenario, on the epistemic construction). The case of two haecceitistically different but qualitatively identical worlds illustrates this: the haecceitistic differences are irrelevant for epistemic purposes.[18]

For every scenario, one world (in the scenario's space of worlds) will be the world *associated* with the scenario. Intuitively, this is the world that will be actual if the scenario obtains. If scenarios are centered worlds, a scenario's associated world will be the scenario stripped of its center. On the epistemic view of scenarios, we can say that (to a first approximation) W will be associated with V when canonical descriptions of V and W are epistemically compatible. (Note that this definition allows that in principle more than one world could be associated with a scenario, if scenarios are relevantly less fine-grained than worlds.[19])

[18] This bears on the issue about intrinsic properties raised by Schroeter (2004). I would like to think that there is a semantically neutral conception of fundamental intrinsic properties, but the framework is not committed to this. If there is no such conception, then one will have to use non-neutral language to fully characterize worlds on the second dimension. The first dimension will be unaffected, however: at worst, a single maximal hypothesis will correspond to an equivalence class of centered worlds (see footnote 9). At most, what is affected is the alignment between epistemic space and subjunctive space: epistemic space will be in this respect smaller than subjunctive space. There will still be a reasonably robust link between epistemic and metaphysical possibility: the resulting position will be what Chalmers (2002a) calls "strong modal rationalism" without "pure modal rationalism" (on this position, conceivability entails possibility, but possibility does not entail conceivability, due to the existence of "open inconceivabilities"). This is an instance of the general point that semantic neutrality is relevant to the alignment between the epistemic and the subjunctive, but is inessential to purely epistemic issues.

[19] For example, if intrinsic properties operate as in the previous note, then two worlds with different but isomorphic distributions of intrinsic qualities may be associated with the same scenario.

Given the association relation between scenarios and worlds, one can define the *diagonal intension* of a sentence's two-dimensional intension. This will be a mapping from scenarios to truth-values, mapping V to the value of the two-dimensional intension at (V, W), where W is associated with V. (If there is more than one such W for the reasons above, it is not hard to see that they will all give the same results.) The diagonal intension of a sentence will straightforwardly be equivalent to its epistemic intension. One can therefore reconstruct an expression's epistemic intension from its two-dimensional intension by diagonalizing, just as one can reconstruct its subjunctive intension by holding fixed the actualized scenario.

It should be clear, however, that this diagonal construction in no sense gives the *definition* of an epistemic intension. Epistemic intensions are defined in purely epistemic terms: they are in no sense derivative on subjunctive notions. The diagonal construction is conceptually much more complex, involving subjunctive evaluation, association of worlds with scenarios. In effect, the relation is akin to that between the functions $f(x) = x^3$, $g(x, y) = x^3 + sin(x - y)$, and $g'(x) = g(x, x)$. Here, g' is the "diagonal" of g, and is the same function as f. But it would obviously be incorrect to hold that f is fundamentally the diagonal of g, or that it is derivative on trigonometric notions. For exactly the same reasons, it is incorrect to hold that an epistemic intension is fundamentally a diagonal intension, or that it is derivative on subjunctive notions.

There is a sense in which the two-dimensional intension represents the full modal structure of an expression, capturing how it behaves under epistemic evaluation, modal evaluation, and combinations of the two. Just as an epistemic intension can be evaluated a priori, a two-dimensional intension can be evaluated a priori. A subjunctive intension cannot be evaluated a priori, but it can be evaluated when the actualized scenario is specified.

We can think of all of these intensions as aspects of the content of a sentence token. A sentence is in no sense ambiguous for having both epistemic intensions and subjunctive intensions; rather, it has a complex semantic value. Different aspects of this semantic value will be relevant to the evaluation of the sentence in difference contexts. In certain epistemic contexts ('it is a priori that S'; 'it might turn out that S'; 'if S is the case, then T is the case'), the epistemic intension of S may play a key role in determining the truth-value of the complex sentence. In subjunctive contexts ('it might have been that S'; 'if it had been that S, it would have been that T'), the subjunctive intension of S may play the most important role. In combined epistemic-subjunctive contexts, truth-value may depend on the two-dimensional intension of S. As usual, there is no need to settle the question of which of these, if any, is *the* meaning or content of an expression.

3.11 The Core Thesis

Let me summarize where things stand with respect to the Core Thesis: that S is a priori iff S has a necessary 1-intension.

First, if S is a priori: then for all W, if D is a canonical description of W, then D implies S. (If S is a priori, D implies S for *any* D.) So S is verified by all W, and has a necessary 1-intension.

Second, if S is not a priori and we take the epistemic approach to scenarios: then ∼S is epistemically possible. Under small assumptions (see 3.4.2), it follows that there is an epistemically complete D such that D implies ∼S. Any epistemically complete sentence describes a scenario, so there is a scenario W that verifies ∼S. So S does not have a necessary 1-intension.

Third, if S is not a priori and we take the metaphysical approach to scenarios: then if Metaphysical Plenitude is true, any epistemically complete D describes a scenario (a centered world), so S does not have a necessary 1-intension. If Metaphysical Plenitude is false, this does not follow: some epistemically possible statements will not be verified by any centered world, so the Core Thesis will be false.

It follows that the Core Thesis is true on the epistemic approach to scenarios, and that it is true on the metaphysical approach iff Metaphysical Plenitude is true. I think there is good reason to hold that Metaphysical Plenitude is true; but even if it is not, we may simply adopt the epistemic approach to scenarios. Either way, the epistemic understanding of two-dimensional semantics plausibly yields an understanding of 1-intensions that satisfies the Core Thesis.

In effect, the epistemic understanding of two-dimensional semantics reconstructs the golden triangle by taking certain epistemic notions as basic and defining certain semantic notions in terms of them, with the aid of modal notions. On the epistemic approach to scenarios, the order of explication is as follows: we take an epistemic notion (such as apriority) as basic, use this to define a modal space (the space of epistemic possibilities), and use this to define corresponding semantic entities (epistemic intensions). On the metaphysical approach to scenarios, we take both an epistemic notion (such as apriority) and a modal notion (metaphysical possibility) as basic, and combine the two to define corresponding semantic entities. On the former approach, the strong connection to the epistemic domain is more or less guaranteed by the construction. On the latter approach, it is grounded in the construction along with the thesis of Metaphysical Plenitude, which articulates a connection between the epistemic and modal domains.

On this approach, the connection between meaning and reason is built in to a large extent by definition. This suggests that we should not make the claim embodied in the golden triangle too strong: a semantic pluralist should accept that there are many other aspects of meaning that are not connected in this way to the epistemic domain. But at the same time, it does not render the analysis trivial. Sense is definitionally connected to cognitive significance, and (subjunctive) modal intensions are definitionally connected to metaphysical possibility, but each of these semantic notions has a powerful role to play. Their cash value is grounded in the phenomena that they help us to analyze. Likewise, in the case of epistemic intensions, the fact that there is a semantic value that bears these connections to the epistemic and modal domains allows us to use semantic and modal tools to play an important role in analyzing the epistemic properties of language and thought.

3.12 Applications

This role for epistemic intensions can be brought out in a number of applications that I will simply summarize here.

(i) *Fregean sense* (see Chalmers 2002b): Because they satisfy the Core Thesis, epistemic intensions also satisfy the Neo-Fregean Thesis: 'A' and 'B' have the same intension iff 'A ≡ B' is a priori. So epistemic intensions behave broadly like a sort of Fregean sense, tied to the rational notion of apriority. There are some differences. First, sentence-level Fregean senses are supposed to be true or false absolutely, but sentence-level epistemic intensions are true or false relative to a speaker and time (witness 'I am hungry now'). Second, apriority is weaker than cognitive insignificance, so epistemic intensions are less fine-grained than Fregean senses. (One might adapt the current framework to yield a more fine-grained sort of epistemic intension, by starting with a less idealized notion of epistemic possibility; see Chalmers (forthcoming).) Nonetheless, epistemic intensions can serve as a broadly Fregean semantic value.

(ii) *Narrow content* (see Chalmers 2002c): One can extend the current framework from language to thought in an obvious way. One can define epistemic intensions for beliefs and thoughts in the manner suggested in 3.8. The result can be seen as a sort of content of thought. It is very plausible that what results is a sort of *narrow* content, such that two physical and phenomenal duplicates will have thoughts with the same epistemic intension. (This narrowness is grounded in the narrowness of deep epistemic possibility: if a thought is epistemically necessary, then the corresponding thought of a physical and phenomenal duplicate will also be epistemically necessary.) This sort of content is much more closely tied to cognition and reasoning than "wide content", and is well-suited to play a central role in explaining behavior.

(iii) *Modes of presentation* (see Chalmers 2002c, section 8): In analyzing the behavior of belief ascription, it is common to appeal to a notion of "mode of presentation", but there is little agreement on what sort of thing a mode of presentation is. Schiffer (1990) suggests that a mode of presentation must satisfy "Frege's constraint": roughly, that one cannot rationally believe and disbelieve something under the same mode of presentation. Because they satisfy the neo-Fregean thesis, epistemic intensions satisfy Frege's constraint perfectly, at least if one invokes an idealized notion of rationality that builds in arbitrary a priori reasoning. So it is natural to suggest that modes of presentation are epistemic intensions. In this way, one can use epistemic intensions to analyze ascriptions of belief.

The current framework is compatible with a number of different proposals that give modes of presentation a role in belief ascription. A naive first account might suggest that 'X believes that S' is true if the subject specified has a belief whose epistemic intension is the epistemic intension of S (for the ascriber), but numerous counterexamples to this claim immediately present themselves.[20] A more plausible account follows the general shape of so-called "hidden-indexical"

[20] Soames (2004) attributes this naive account of belief ascriptions to "strong two-dimensionalism", and criticizes the resulting view. These criticisms have no force against the view of belief ascriptions laid out in Chalmers (1995) and Chalmers (2002c).

accounts (Schiffer 1990). On such an account, at a first approximation, 'X believes that S' will be true if the subject specified has a belief whose subjunctive intension is that of S, and which has an appropriate epistemic intension, where the range of appropriate epistemic intensions may be contextually determined.

(iv) *Indicative conditionals* (see Chalmers 1998 and Weatherson 2001): One can use epistemic intensions to give a semantics for indicative conditionals that parallels in certain respects the common possible-worlds semantics for subjunctive conditionals. As a first approximation, one can suggest that an indicative conditional 'If S, then T' uttered by a subject is correct if the epistemically closest scenario that verifies S also verifies T, where epistemic closeness will be defined in terms of the beliefs or knowledge of the subject. (Weatherson (2001) pursues a closely related idea.)

(v) *Conceivability and possibility* (see Chalmers 2002a): The Core Thesis makes possible a certain sort of move from conceivability to possibility. If we say that S is conceivable when its negation is not a priori, then when S is conceivable, there will be a scenario verifying S. If we understand scenarios as possible worlds and if Metaphysical Plenitude is true, then when S is conceivable, there will be a centered possible world verifying S. This makes it possible to move from epistemic premises to modal conclusions, as is often done. Of course it is possible to embrace the current framework while rejecting Metaphysical Plenitude and so rejecting the relevant move from conceivability to possibility. But the current framework at least shows how a certain sort of link between conceivability and possibility is tenable in light of the Kripkean phenomena that are often thought to be the greatest threat to such a connection.

4. Epistemic Intensions and Contextual Intensions

We have seen that there are two very different ways of understanding two-dimensional semantics: the epistemic understanding and the contextual understanding. On the epistemic understanding, 1-intensions are constitutively tied to the epistemic domain and satisfy the Core Thesis. On the contextual understanding, 1-intensions are not constitutively tied to the epistemic domain and do not satisfy the Core Thesis.

It is useful to examine the relationship between the two in somewhat more depth. First I will examine how the epistemic understanding deals with the problems that arise for the contextual understanding. Then I will examine the resemblance of certain contextual intensions to epistemic intensions.

4.1 Problem cases

The first main problem area for contextual intensions involved sentences such as 'A sentence token exists', which are a posteriori, but have a necessary contextual intension. These problems arose because contextual intensions require a token of the evaluated expression in the evaluated world. There is no such requirement for epistemic intensions, so the problem does not arise.

For example, there will be many language-free scenarios: there are many centered worlds with no sentence tokens, and there are many epistemically possible hypotheses according to which there are no sentence tokens. If D is a canonical description of such a scenario, D will verify 'There are no sentence tokens'. Intuitively, if we consider such a scenario W as actual, we can say that *if* W is actual, then there are no sentence tokens. So the epistemic intension of 'A sentence token exists' will be contingent, as required.

The same goes for 'words exist', and something similar applies to 'I am uttering now'. In the latter case, there will be many centered worlds in which the subject at the center is not uttering, and there will be many epistemically possible hypotheses (for me) under which I am not uttering. If D is a canonical description of such a scenario, D will verify 'I am not uttering now'. So this expression will also have a contingent epistemic intension. The same applies even to 'I am thinking now'.

'I exist' is a slightly trickier case. If 'I exist' is a priori, there is no problem. If 'I exist' is a posteriori (as I think is the case), then there will be various epistemically possible hypotheses for me under which I do not exist: for example, a hypothesis under which nothing exists (which is arguably itself not ruled out a priori). So on the epistemic view, there will be corresponding scenarios that verify 'I do not exist', and 'I exist' will have a contingent epistemic intension, as required.

On the world-based view, there is a worry: one might think that any centered world will verify 'I exist', since there is always a subject at the center. This raises a subtlety. In the general case, centering is *optional*: on the world-based view, the space of scenarios contains worlds without a marked subject and time, and perhaps worlds with only a marked subject or only a marked time. A world without a marked subject will then verify 'I do not exist'. The exact choices here will depend on exactly which indexical claims one holds to be a priori, but it should be possible to arrange things so that there is a verifying centered world for every epistemically possible claim.

In any case, we see that problems that arise due to the required presence of a token do not arise here. At most there are problems due to the required presence of a *subject* in a centered world; but these will not arise on the epistemic view, and can be dealt with reasonably straightforwardly on the world-based view. So the epistemic understanding does not suffer from the problems of the contextual understanding here.

Another problem, at least for orthographic contextual intensions, concerned worlds where the subject at the center uses 'bachelor' to mean something different, such as horse, so that the 1-intension picks out horse there, which is not the desired result. Again, this problem will not arise for epistemic intensions. In general, to evaluate the epistemic intension of 'bachelor' at a scenario, the presence or absence of tokens of 'bachelor' in that scenario will be irrelevant (with one qualification to be outlined shortly). What the epistemic intension of 'bachelor' picks out in a Steel Earth scenario will depend on a number of other factors, especially the appearance and behavior of substances located around the center of the scenario, but there is no danger that it will pick out steel.

Note that this analysis requires that "My term 'bachelor' means bachelor" and similar claims are not a priori. If such a claim was a priori, then because a canonical description of the Steel Earth scenario will contain something like "My term

'bachelor' means horse'', the scenario would verify 'bachelors are horses', which is the wrong result. But it is independently plausible that these claims are not a priori—at least *if* '' 'bachelor' '' is understood in purely orthographic terms. It is a posteriori that the string has any meaning at all, and it is a posteriori that it means what it does. If '' 'bachelor' '' is understood in partly semantic terms, so that it is constitutively tied to a given meaning for 'bachelor', then the claim in question may be a priori; but this is not a problem, since in this sense the Steel Earth scenario will not verify ''My term 'bachelor' refers to horse''. For more on this matter, see Chalmers (2002a) and Yablo (2002).

Note also that there is one case where evaluating an expression's epistemic intension may turn on the presence of tokens of that expression in a world: expressions used *deferentially*. It may be that Leverrier's wife uses 'Neptune' to (rigidly) pick out whatever her husband refers to as 'Neptune'. If so, then in a given scenario, the epistemic intension of 'Neptune' will pick out roughly the referent of her husband's term 'Neptune' in that scenario (abstracting away from issues about the epistemic intension of 'my husband', etc.). In this case, something like ''If Neptune exists, my husband refers to it as 'Neptune' '' will be a priori for her. Something similar to this will apply to other terms used deferentially, such as a non-expert's use of 'arthritis', although the details may be less clean. But here, we get only the results that would be expected. For example, if I use 'water' wholly deferentially, then if I consider as actual a Steel Earth scenario where those around me use 'water' for steel, then this scenario verifies 'Water is steel' for me. This seems correct: for a deferential user, although perhaps not for a nondeferential user, 'Water is steel' expresses an epistemically possible thought.

(Note that even in deferential cases, evaluation turns on the referent of *others'* use of the expression. It may be that evaluation could also turn on one's own past use of the expression; but it cannot happen that evaluation will turn on the referent of one's own current use of the expression, since such a circular criterion cannot secure a referent. (I set aside pathological nonreferring cases, such as 'the referent of this expression'.) So even in a strongly deferential case, the epistemic intension of 'water' will not turn on the referent of a use of 'water' at the very center of a scenario.)

There is also the Twin Earth case, where Twin Oscar uses 'water' to refer to XYZ. This was a problem for linguistic and semantic contextual intensions, since these are arguably not defined at such a world, whereas we would like the 1-intension of Oscar's term to return XYZ at this world. Again, the epistemic framework handles this unproblematically. The epistemic intension of 'water' returns XYZ at this world, *not* because Twin Oscar's term 'water' refers to XYZ (Twin Oscar's term is irrelevant), but because the scenario verifies the claim that XYZ has a certain appearance, behavior, relation to oneself, and so on, which in turn verifies 'Water is XYZ'.

Finally, there was the problem of Fregean typing. It seemed that in order for contextual intensions to give roughly Fregean results, then one had to classify expression tokens under some sort of Fregean type. For a semantic contextual intension to give the right results, for example, one needed to appeal to some sort of prior Fregean semantic notion, which is unhelpful in the current context. No such problem applies

to epistemic intensions. Because these intensions do not rely on tokens of the same type being present within scenarios, there is no need to isolate the common type under which these tokens fall. All one needs is the expression token itself, and its epistemic properties. This approach may *ground* an account of a sort of Fregean semantic value, but it need not presuppose any such account.

These advantages of the epistemic account over the contextual account are all grounded in the fact that the contextual understanding is an essentially metalinguistic understanding, while the epistemic understanding is not. The contextual understanding concerns content that an expression might have had; but the epistemic understanding reveals aspects of the content that it has. Everything is grounded in certain first-order epistemic claims, which we use as tools to reveal an expression's content, just as in the familiar modal case, various first-order subjunctive claims are used as tools to reveal an expression's content. As before, the cases are parallel.

4.2 Semantic contextual intensions

We saw earlier that some versions of a semantic contextual intension presupposed a quasi-Fregean notion of content. We can now turn the picture the other way around, using the quasi-Fregean notion of content developed here to ground a semantic contextual intension.

Let us say that an *epistemic contextual intension* of an expression is the semantic contextual intension that derives from the use of epistemic intensions as the relevant semantic value. The epistemic intension of an expression token is a function from centered worlds to extensions, defined at worlds that have a token at the center with the same epistemic intension as the original token, and returning the extension of that token in that world.

It is easy to see that at the worlds where it is defined, an expression's epistemic contextual intension yields the same extension as its epistemic intension. If W is a centered world containing a token S′ with the same epistemic intension as the original token S: let E be the extension of S′. Then the epistemic contextual intension of S returns E at W. Further, the epistemic intension of S′ (on the world-based view of scenarios) returns E at W, since W is actualized at S′. By identity of epistemic intensions, the epistemic intension of S also returns E at W. So S's epistemic contextual intension and epistemic intension are coextensive at W. Something similar applies on the epistemic view of scenarios, if we invoke the scenario corresponding to the centered world W.

So an expression's epistemic contextual intension is a restriction of the term's epistemic intension. For this reason, it will give appropriate quasi-Fregean results in many cases. It will not satisfy the Core Thesis: it will have the usual problems with 'A sentence token exists' and other metalinguistic claims, as it will not be defined at scenarios without the token at the center, or where the token has a different content. But it will be reasonably close for many purposes.

One could also define a epistemic version of the *cognitive* contextual intension of a token, defined at all worlds centered on a *concept* or *thought* with the same epistemic intension as the token, and returning the extension of that concept or thought.

This would again be a restriction of the expression's epistemic intension, but it would be less of a restriction, since it would not require a linguistic token in the evaluated world. The Core Thesis will still be false due to various metacognitive claims and the like; but it will not be far off. I think this last notion is the best approximation that a contextual intension can yield to a quasi-Fregean content that satisfies the Core Thesis. It is clear, however, that this notion is essentially derivative of that of an epistemic intension.

This way of seeing things also helps to explain why some other contextual intensions give approximately Fregean results. It may be that there are various other features of type F of a subject or a token that at least correlate with an epistemic intension to some degree. We can then set up an F-based contextual intension, defined at worlds centered on a subject or token with the same F features as the original, and returning the extension of the relevant token. Then in each such world, the token at the center will have at least approximately the same epistemic intension as the original token, and so in most cases will return the same or similar extension at that world. So the F-based contextual intension will approximate the behavior of a restriction of the original epistemic intension.

This applies especially to some cognitive contextual intensions. It may be that possession of a concept with a given epistemic intension is itself determined by features such as a concept's cognitive role and/or associated phenomenology, or more deeply by the subject's physical state, or functional state, or physical/phenomenal state. To know exactly which features are crucial would require a solution to the problem of intentionality, which is not yet available. But one can say: *insofar* as epistemic intensions are determined by features such as cognitive role or physical/phenomenal state, then corresponding contextual intensions (here, cognitive-role contextual intensions or physical/phenomenal contextual intensions) will be restrictions of the original epistemic intension, and so will behave in a quasi-Fregean manner. Again, however, the epistemic intension is the more fundamental notion of content.

4.3 Linguistic contextual intensions

We saw earlier (in Section 2.2) that for some expressions, a linguistic contextual intension behaves in a quasi-Fregean manner. We are now in a position to see why this is.

For some expressions, their epistemic intension is part of (or determined by) their linguistic meaning. That is, some linguistic expression types are such that every token of that type has the same epistemic intension. As noted in Section 3.7, something like this appears to apply to some pure indexicals, such as 'I', 'now', and 'today', to some descriptive terms, such as 'circular', and to some descriptive names, such as 'Jack the Ripper'.

When an expression token's epistemic intension is part of its linguistic meaning, then the token's linguistic contextual intension will be a restriction of its epistemic intension. This can be seen by the same sort of reason as in the previous section. Or one can simply apply the point there directly: if any token of S's linguistic type has the same epistemic intension, then S's linguistic contextual intension will be a restriction of its epistemic contextual intension, which is a restriction of its epistemic intension.

It follows that in cases such as 'I', 'now', 'circular', and 'Jack the Ripper', the terms' linguistic contextual intensions will be quasi-Fregean. They will not satisfy the Core Thesis because of the restriction to worlds containing relevant tokens, but they will be reasonably close. This explains the phenomenon noted in Section 2.2: the quasi-Fregean behavior is a direct consequence of the fact that for these tokens, epistemic intension is an aspect of linguistic meaning. Once again, a contextual intension is interesting largely because of the degree to which it approximates an epistemic intension.

5. Other Varieties of Two-dimensionalism

With this analysis of the contextual and epistemic understandings on the table, we are now in a position to turn to existing two-dimensional proposals to see how they fit into this analysis, and to use this analysis to help understand their foundations. I should say at the start that although I will occasionally criticize these approaches and argue that the approach I have recommended has certain advantages, any advantages are due largely to building on the insights that these approaches embody.

5.1 Stalnaker's diagonal

The diagonal proposition of Stalnaker (1978) is characterized as follows. We start with an understanding of propositions as sets of possible worlds, and with the idea that any utterance has a proposition as its content. (This propositional content coincides roughly with what I have called a subjunctive intension.) We can then say: the utterance could have had different propositional content. So there are worlds where the utterance has different propositional content. This allows us to define an utterance's *propositional concept*, which is a function from possible worlds to propositions, defined at any world containing the utterance, returning the propositional content of the utterance at a world. We can then define the utterance's *diagonal proposition* as the set of worlds such that the utterance's propositional concept, evaluated at that world, yields a proposition that is true at that world.

As defined here, a diagonal proposition is much like a token-reflexive contextual intension. There are minor differences. A token-reflexive contextual intension was defined directly in terms of what an utterance's truth-value would be at a world, rather than in terms whether the proposition it expresses would be true at that world, but it is clear that within the propositional framework, these yield the same results. A diagonal proposition is a set of possible worlds or equivalently a function from worlds to truth-values, whereas a token-reflexive proposition was a function from centered worlds to truth-values. But again, that is a minor difference: one can translate between the relevant worlds and centered worlds either by "marking" the location of the token as a center, or by removing the marked center from the location of the token. (Token-reflexive contextual intensions uniquely do not need a center to specify the relevant token, since the token is independently identified by transworld identity with the original token.) So diagonal propositions are equivalent to token-reflexive contextual intensions.

The behavior of a token-reflexive contextual intension is not clear unless we know which properties are essential to a token and which properties are inessential. It is clear that Stalnaker holds that a token's semantic properties are not essential to it, since he holds that a token could have had different semantic content. It seems plausible that on his picture, a token's orthographic properties are essential to it; at least, in all examples, a token's orthographic properties are held constant across its possible occurrences, so I will assume this in what follows. It is not entirely clear which other properties are essential (language? probably not; speaker? maybe), but we need not settle that issue for now.

From what we have said here, it appears that a token's token-reflexive contextual intension will be a restriction of its orthographic contextual intension, restricted to cases where the orthographically identical token *is* the original token. One might think that its general behavior will be very much like that of an orthographic contextual intension: for my utterance 'Water is H_2O', there will be worlds at which it means that steel is orange; if so, its token-reflexive contextual intension will be defined and presumably false there.

However, this sort of understanding is inconsistent with a central point in Stalnaker's 1978 paper, where he says that for a sentence to have a necessary diagonal proposition is for it to be an a priori truth. He also says that an official's utterance of 'this bar is one meter long' could not have expressed a false proposition. It is unclear how this can be justified. Why could not the utterance have meant something like "that boat is two miles long", and been false? It is natural to suppose that Stalnaker is holding fixed some intuitive sort of meaning and content (thus yielding something that behaves like a semantic/orthographic contextual intension). But it is not clear what aspect this could be: the only content that his account officially recognizes is propositional content, which is explicitly held to vary with possible occurrences of an utterance; and even if there were some further aspect of content, it is unclear why this sort of content should be essential to an utterance and the other sort inessential. Alternatively, it may be that Stalnaker is assuming that some sort of cognitive factor (for example, associated cognitive role?) is essential to an utterance (thus yielding something that behaves like a cognitive/orthographic contextual intension), but this is nowhere specified. In this article, the connection between apriority and diagonal propositions appears to be ungrounded.

In later work, Stalnaker does not repeat the claim about apriority, and he allows a much wider range of behavior for an utterance across possible worlds. For example, in Stalnaker (1999) he allows that there are worlds where 'Julius', used in the actual world as a descriptive name for the inventor of the zip, is used instead for the inventor of bifocals. And in Stalnaker (2001) he allows that there are worlds where our word 'tiger' refers to pieces of furniture. In both of these articles he explicitly denies that necessity of diagonal proposition corresponds to apriority, as seems reasonable. In effect, this diagonal proposition behaves very much like an orthographic contextual intension, or a straightforward restriction thereof.

As such, a diagonal proposition is clearly useful. For example, Stalnaker often uses diagonal propositions to model situations of communication, in which a hearer hears an utterance but is unsure what the speaker meant by it, or has false beliefs about

what the speaker meant by it. It seems clear that this sort of metalinguistic use requires something quite different from a quasi-Fregean notion, so this is reasonable. It is less clear that diagonal propositions are useful for addressing matters of cognitive significance, rational inference, apriority, and the like, especially when divorced from issues about communication. Stalnaker sometimes suggests this sort of use, but I think the grounds here are weaker.

For example, in a recent paper (2001), Stalnaker holds that the "metasemantic" framework with diagonal propositions can "provide an explanation for the phenomena that Kripke's work brought to light"—where this phenomenon is the distinctive behavior of the class of a posteriori necessities such as 'Hesperus is Phosphorus', 'Water is H_2O', and so on. This is a surprising claim. What is most distinctive about these phenomena are the *differences* with standard necessities such as 'All bachelors are unmarried', '$2 + 2 = 4$', and so on. If diagonal propositions function as Stalnaker (1978) suggests, the distinction would be straightforwardly represented by the fact that the second class have necessary diagonal propositions and the first do not. But on the metasemantic understanding, there seems to be no way to draw the distinction using diagonal propositions alone. For both sorts of necessity, there will be many worlds at which the diagonal is false, and there are no clear patterns that are distinctive to the first class. So it is not clear how the "explanation" is supposed to work.

In a brief ensuing discussion of 'Hesperus is Phosphorus', Stalnaker appeals to the fact that there is a world where the diagonal proposition is false. But clearly this holds equally for 'all bachelors are unmarried'. One might find some differences if one focuses on a *restriction* of the diagonal proposition. It is notable that in the worlds Stalnaker discusses, 'Hesperus' and the like appear to be used with the same reference-fixing intentions as the original term, for example. It may be that under this sort of restriction, the two sorts of necessities behave differently: there are counterexamples to the diagonal for one class but not the other. But the restriction is doing all the work: in effect, it invokes something more like a cognitive contextual intension, rather than a diagonal proposition per se. Such intensions may be able to model the Kripkean distinction at least approximately, if imperfectly. Stalnaker appears to use similar tacit restrictions in some other cases; in all these cases, it seems that in effect a restricted contextual intension does the explanatory work.

The explanatory power of restricted contextual intensions here itself plausibly derives from that of epistemic intensions. Epistemic intensions handle these phenomena straightforwardly: for Kripkean necessities, there is a falsifying scenario, and for standard necessities there is not. For reasons we have seen, certain restricted contextual intensions approximate epistemic intensions, and so approximate this behavior (with some exceptions). So it is plausible that the usefulness of diagonal propositions in this context derives indirectly from the usefulness of epistemic intensions.

Stalnaker contrasts his "metasemantic" version of the framework with "semantic" versions, on which 1-intensions are an aspect of semantic content, and suggest that the apparent attractions of the latter in explaining these phenomena derive from the attractions of the former. The above suggests that this is not quite right: the attractions of Stalnaker's version of the framework in this domain derive from the

attractions of the epistemic version. As for whether the epistemic understanding is itself a "semantic" understanding: this matter depends on what is meant by "semantic". Stalnaker mostly uses the term to contrast with "metasemantic", indicating an aspect of first-order content rather than a metalinguistic notion: in this sense, epistemic intensions are semantic. Stalnaker also sometimes uses the term to indicate those aspects of content that are built into linguistic expression types, rather than varying across tokens: in this sense, epistemic intensions are not semantic. Stalnaker appears to assume that his opponent's framework is semantic in both these senses;[21] but these are very different distinctions. Epistemic intensions need not be built into linguistic meaning to be a sort of first-order content that does explanatory work.

In any case, I think it is clear that diagonal propositions and epistemic intensions both have useful roles to play. Diagonal propositions are best suited to analyzing matters of context-dependence, and epistemic intensions are best suited to analyzing the epistemic domain.

5.2 Kaplan's character

Kaplan's notion of character is set out as follows. We assume a prior notion of the proposition expressed by an utterance: such a proposition is something in the vicinity of a 2-intension, although it may be a singular proposition instead. For some linguistic expression types (e.g. 'I'), utterances of the same type can express different propositions in different contexts. The character of an expression type is a function from contexts to propositions, returning the proposition that an utterance would express in a given context.

At first glance, it may seem that character is much like a linguistic contextual intension. There are some superficial differences. For example, Kaplan's contexts are not quite centered worlds, but they include an "actual-world" and a few other parameters (speaker, time, etc), so they can be modeled by centered worlds. Also, character is a function from contexts to propositions, not to extensions. But one can diagonalize character by evaluating the proposition associated with a given context in the world of that context, yielding an associated function from contexts to extensions.

In many cases, (diagonalized) character behaves quite like a linguistic contextual intension. We have seen that the linguistic contextual intension of indexicals such as 'I', 'now', and 'today' pick out the speaker, time, day (and so on) of the center of all worlds at which they are defined. The same is true for (diagonalized) character, on Kaplan's account. At the same time, we have seen that the linguistic contextual intension of a name arguably picks out the same individual at all worlds where it is defined. Again, the same applies to character, on Kaplan's account.

[21] The only opponent that Stalnaker cites is Chalmers (1996). I note that Chalmers (1996) explicitly leaves open (p. 58) the question of whether different speakers might associate different 1-intensions with the same word. Stalnaker also argues that the framework cannot yield an account of the a priori; I agree, and have not claimed that it can.

One case where the two apparently behave differently is for demonstratives such as 'that'.[22] If I use 'that' intending to refer to an object in front of me, then its character will pick out (roughly) an object in front of the speaker in all contexts. But the linguistic contextual intension will not: what it picks out in a context will depend on the underlying demonstration or intention of a speaker in that context. But in a way, this is the exception that proves the rule. In Kaplan's formal analysis, he stipulates that different uses of 'that' (roughly, those corresponding to different demonstrations) are tokens of *different* words: 'that_1', 'that_2', and so on. Under this stipulation, it is plausible that a linguistic contextual intension for one of these instances of 'that' will behave as characterized above.

However, there are aspects of Kaplan's discussion that make it clear that character is fundamentally different from a linguistic contextual intension. Kaplan stresses that when we evaluate a sentence's character in a context, we do not evaluate an *utterance* of that sentence within the context. Rather, we evaluate an *occurrence* of the sentence at the context. An occurrence is in effect an ordered pair of a sentence and a context. And crucially, the context need not itself contain an utterance of the sentence. In effect, this is to allow that the character of an expression can be evaluated directly at a centered world, whether or not the world contains a token of the original expression.

Kaplan's reason for doing this are largely tied to his desire for a *logic* of demonstratives. He suggested that arguments involving demonstratives should be *valid* in virtue of their character: that is, a conclusion should follow from premises in virtue of an appropriate relation among their characters. But if the character of a claim were restricted to contexts containing an utterance of that claim, then each premise and the conclusion would be defined across different contexts, so their characters could not stand in the right sort of relation. He also says that there are sentences that express a truth in certain contexts, but in no contexts in which they are uttered: for example, 'I say nothing'. If so, contexts cannot be required to contain a token of the relevant utterance.

For these purposes, it is natural to suggest that character should be something more like an epistemic intension. Validity, at least as Kaplan uses it here, is a deeply epistemic notion, tied to apriority and to rationally compelling inferences (in Kaplan's discussion, it is clear that validity is not tied constitutively to necessity). The sort of intension that is *constitutively* tied to validity and to apriority is an epistemic intension. Similarly, for the intension of 'I say nothing' to be false in the relevant contexts, the best candidate is something like an epistemic intension.

It is difficult to adjudicate what Kaplan intends, however, since he never specifies how to evaluate an expression's character in a context. He simply stipulates that expressions have a character associated with them, and then discusses the character's properties. He does say on some occasions that character picks out what the expression would pick out if uttered in that context, but he retracts this because of the point about occurrences. (It presumably remains the case that *when* a context contains the right sort of utterance, character returns what the utterance picks out).

[22] I use "demonstrative" for expressions such as 'that', 'he', and 'you', while using "indexical" for expressions such as 'I', 'here', and 'now'.

He also says (505) that character is set by linguistic conventions and determines the content in a context, and he suggests that character is determined by a demonstration (526–7) or a directing intention (587–8). But nothing here tells us how to evaluate character in contexts not containing the utterance. In some cases the matter seems reasonably straightforward: 'I' picks out the marked subject in a context, 'you' picks out a marked addressee. But there seems to be no general principle here for assigning an evaluation function to an expression type.

(Another complication is that it is not entirely clear what is built into the relevant context. On a couple of occasions (528, 588) Kaplan entertains the idea that a context explicitly contains a parameter for a demonstratum, which serves as referent for a demonstrative. If this is done, it renders the question about how to evaluate character trivial, but at the cost of trivializing many other aspects of the framework. It also removes any special role for demonstrations and directing intentions in contextual evaluation, and removes the deep connection with cognitive significance. Partly for these reasons, and partly because it eliminates the connection between demonstratives and indexicals, this seems not to be Kaplan's considered view. The alternative is a view on which the referents of demonstratives are not explicitly specified within a context, but instead are picked out by a directing intention.)

One way to address the question is to ask: is it *constitutive* of character that validity and apriority are governed by character? Or is this merely a feature that character turns out to have? It seems that it cannot be constitutive, for the obvious reason that in the case of proper names, validity and apriority come apart from character. But now the question arises: what justifies Kaplan's claim that the character of indexicals and demonstratives must be logically well-behaved? In effect, it is this claim that determines his treatment of occurrences. If character were definitionally connected to validity, the claim would be reasonable; but character is not definitionally connected to validity. If character is independently grounded, it seems that one might equally say: character is reasonably close to reflecting validity and the like, but unfortunately the correspondence is imperfect, even for indexicals and demonstratives.

Perhaps the most likely diagnosis is the following. The initial notion of character is not constitutively connected to the domain of validity and apriority. (Perhaps it is something like a linguistic contextual intension.) But at least in the case of indexicals and demonstratives, this notion of character turns out to come very close to reflecting this domain. It turns out that a slight modification makes the correspondence precise, so at least in this case we adopt the modified notion. The resulting notion appears to be something quite close to an epistemic intension. It is not exactly an epistemic intension, for example because of the use of further parameters in a context. But it seems to behave in a quite similar way.

This raises the question: why not do the same for names? If character is to be connected to apriority, why not understand the character of a name so that it behaves something like an epistemic intension? The initial answer is that Kaplan thinks that names do not behave this way: their contents are essential to them, so they do not pick out different contents in different contexts. In discussing the matter, Kaplan notes especially (562) that occurrences of 'Aristotle' that refer to different people are different words. One might respond that this would be relevant if we were defining

the *contextual* intension of names, but we are now dealing with a modified notion. The fact that a name has its referent essentially (and the point about 'Aristotle') is compatible with its *epistemic* intension picking out different referents in different contexts.[23] But the crucial point may in fact be something different: character is supposed to be a sort of linguistic meaning, but the names as linguistic types do not have epistemic intensions (at best, epistemic intensions vary between tokens).[24] So character cannot be epistemic intension.

Still, an obvious response is that the same holds for demonstratives. Kaplan's formal move of stipulating that different tokens of a demonstratives are different words is clearly something of a convenient trick: the force of this move is to suggest that character need not really be associated with a linguistic type, but with a token.[25] If so, then we could say the same for names, perhaps making a similar stipulation, or perhaps not. It is not clear exactly how the cases are relevantly different. One suggestion is that different tokens of a name seem to be more closely tied together than different tokens of a demonstrative, with some sort of associated assumption of communication, agreement and disagreement, and so on. If so, then assigning all these name tokens a different linguistic type might be even more counterintuitive than for demonstratives. But it is not clear exactly what the rules are here. One should arguably take the real moral of the demonstrative case to be that character is fundamentally a property of tokens rather than linguistic types, in which case it is no longer obvious that names must have trivial character.

In any case, my best guess as to what constitutes character is the following: character is something like an epistemic intension in cases where it is reasonable

[23] I have not mentioned Kaplan's 'Fregean theory of demonstrations', according to which a demonstration does not have its referent essentially. It seems that this point would be highly relevant to a contextual understanding of demonstratives, but it is not so clear that it is required for an epistemic understanding.

[24] This sort of point about the difference in cognitive significance between different tokens of a name is never mentioned explicitly in Kaplan's article, but it may be playing a role implicitly in his claim that names do not have nontrivial character semantically associated with them. A useful diagnostic question would be whether descriptive names (if there are any), such as 'Jack the Ripper', can have nontrivial character. If yes, then variability is plausibly the key reason that standard names have trivial character. If no, then essentiality of referent is plausibly the key reason.

[25] See Braun (1996) on this topic. Braun notes that Kaplan sometimes adopts the informal strategy of taking 'that' to be a single word (type), associating character not with the word but with a word-plus-demonstration pair. This raises the question: since the relevant demonstrations (especially according to the later Kaplan) are a sort of intention, why not analogously associate character with a name-plus-intention pair? Then one will in effect have character for utterances of names.

One might even note: for a use of a name, there can be a directly linked demonstrative. This can happen with an anaphoric use in 'John . . . he . . .', or better, with a simple non-anaphoric 'he' backed by the intention to refer to John (perhaps this can be a mild counterfactual variant on the original utterance). It does not seem entirely unreasonable to say that this last demonstrative has nontrivial character. If so, one could use this character to motivate a nontrivial character in the vicinity of any token of 'John'. If not, then it seems to follow that the Fregean theory of demonstrations is false in at least some cases, and one wants to know more about the rules for associating character with demonstratives and demonstrations.

(perhaps at a stretch) to assign an epistemic intension to a linguistic type. If not, because of variability of epistemic properties between tokens, then character is something else, perhaps more like a contextual intension, or perhaps stipulated to be a function that returns an expression's actual content (whatever it is) at all contexts. Alternatively, if character really is something like "epistemic intension insofar as it is associated with a linguistic type", one might equally simply say that names have no character, rather than saying that they have constant character. That is not Kaplan's official view, but it does not seem wholly contrary to the spirit of his discussion.

There are a couple of interesting diagnostic cases. First, what is the character of a descriptive name, such as 'Jack the Ripper'? Second, what is the character of a context-sensitive predicate such as 'heavy'? Kaplan is silent about these matters, but the behavior of character in these cases might help decide just how the notion of character is grounded.

The distinction between the contextual and epistemic understandings also helps bear on a recent controversy about occurrences. Garcia-Carpintero (1998) argues for the superiority of a Reichenbachian token-reflexive account of indexicals over an account that relies on Kaplan's occurrences. In effect, he suggests that a sort of token-reflexive contextual intension (requiring the token in a context) is truer to the data than an account that does not require tokens. As part of his argument, he denies that there is any reasonable intuition that there are contexts in which 'I am not uttering now' is true. Our discussion makes it possible to render a split verdict here. On a contextual understanding of evaluation in contexts, there is no such intuition. But on an epistemic understanding, there is such an intuition. The intuition, I think, is that 'I am not uttering now' is not false a priori, so that there are epistemic possibilities in which it is true. These epistemic possibilities are scenarios in which the subject at the center is not uttering. The difference between Kaplan's and Reichenbach's frameworks may then be grounded in the fact that Kaplan's semantic value for an indexical is constitutively tied to its epistemic properties, while Reichenbach's is tied to its contextual properties.[26]

In any case, it seems plausible that there are elements of both the contextual understanding and the epistemic understanding in Kaplan's account, not always disentangled. Perhaps character is fundamentally an extension of a linguistic contextual intension; perhaps it is fundamentally a sort of epistemic intension; or perhaps there is no fact of the matter. But it is clear that much of the value of character in the case of demonstratives comes from the fact that in this case, character behaves much as an epistemic intension does (whereas in the case of names, it does not). It does not seem unreasonable to hold that character is useful for

[26] Note, though, that on the framework I have suggested, epistemic intensions are fundamentally assigned to utterances, not to occurrences. Occurrences play a role in that epistemic intensions are *evaluated* at scenarios that need not contain the relevant utterance. This suggests that there are two quite distinct issues dividing the Reichenbachian and the Kaplanian: the issue of whether semantic values should be assigned to utterances or occurrences, and the issue of whether these semantic values can be evaluated in worlds (or contexts) in which the utterance is absent.

epistemic purposes precisely to the extent that it approximates or coincides with an epistemic intension.

5.3 Evans' deep necessity

In addressing Kripke's problems of the contingent a priori, Evans (1979) focuses on the case of descriptive names. He introduces the descriptive name 'Julius', whose referent is fixed as being whoever invented the zip. Then 'Julius invented the zip' seems to be a priori. In analyzing the case, Evans distinguishes between two sorts of necessity: "deep necessity" and "superficial necessity". Instances of the "contingent a priori" (such as 'Julius invented the zip') are superficially contingent but deeply necessary; instances of the "necessary a posteriori" are superficially necessary but deeply contingent. Evans says that whether a statement is deeply necessary or contingent depends on what makes it true; and whether it is superficially contingent depends on how it embeds under modal operators.

Superficial necessity is defined as follows. A sentence Q is superficially contingent if '$\Diamond\sim$Q' is true, or equivalently, if there is some world W where Q is not true$_W$. Here, the possibility operator is clearly subjunctive possibility ("it might have been that"), and the possible-worlds evaluation is clearly subjunctive counterfactual evaluation of the Kripkean sort. So superficial necessity is a second-dimensional notion: S is superficially necessary when it has a necessary 2-intension or subjunctive intension.

Deep necessity and contingency are characterized in the following passage toward the end of Evans' article:

> We have the idea of a state of affairs, or a set of state of affairs, determined by the content of a statement as rendering it true, so that one who understands the sentence and knows it to be true, thereby knows that such a verifying state of affairs exists. A deeply contingent statement is one for which there is no guarantee that there exists a verifying state of affairs. If a deeply contingent statement is true, there will exist some state of affairs of which we can say both that if it had not existed the sentence would not have been true, and that it might not have existed. The truth of the sentence will thus depend on some contingent feature of reality. (Evans 1979, 185)

This passage has a strong epistemic element in the first half; and a strong contextual element in the second half. To understand these we need to examine the discussion earlier in Evans' article.

Evans introduced the notion of the *content* of a sentence earlier as capturing an epistemic element. Evans says that when two sentences have the same content, they are *epistemically equivalent*: a person who understands both cannot believe what one says and disbelieve what the other says. Evans makes a distinction between the content of a sentence and the *proposition* expressed by a sentence, which is a function from possible worlds to truth-values of the sort associated with the modal contexts of superficial necessity. He notes that two sentences that express the same proposition can have different contents, and argues that two sentences with the same content can express different propositions: e.g. 'Julius is F' and 'The inventor of the zip is F'.

Evans holds that there is a notion of "making a sentence true" that is tied directly to content (which he distinguishes from an alternative sense tied to proposition expressed). He says:

> . . . if two sentences are epistemically equivalent, they are verified by exactly the same state of affairs, and what one believes in understanding the sentence and accepting it as true is precisely that some verifying state of affairs obtains. On this conception, the same set of states of affairs makes the sentence 'Julius is F' true as makes the sentence 'The inventor of the zip is F' true. If x, y, z, . . . is a list of all objects, then any member of the set {x's being the inventor of the zip & x's being F; y's being the inventor of the zip and y's being F; z's being the inventor of the zip and z's being F . . . } will suffice to make the sentence true. (Evans 1979, 180)

On this conception, making a sentence true, at least in the case of a descriptive name, seems to involve something like satisfying its epistemic intension. In the sense of 'verify' that I tied to epistemic evaluation, 'Julius invented the zip' will be verified precisely when the conditions that Evans suggests for "making the sentence true" obtains. The claim that epistemic equivalence entails verification (in Evans' sense) by the same states of affairs also suggests a tie to epistemic intension, and suggests a link between the two notions of 'verification'. If deep necessity is tied to 'making true' in this sense, then at least in the case of descriptive names it seems to be a sort of necessity of epistemic intension.

Evans also characterizes this notion of 'making true' in alternative terms:

> But there is an ineliminable modal element in the notion of what makes a sentence true. For what can it mean to say that any one of a set of states of affairs renders a sentence true, other than to say that, if any one of them obtains, the sentence will be true, and if any of them *had* obtained, the sentence *would have been* true.

This characterization has a more contextual flavor: "making true" is characterized in terms of a metalinguistic subjunctive about truth-values the sentence *could* have had. This suggests something like a linguistic contextual intension. But such an understanding will yield quite different results from the understanding above. For example, if 'L' is a descriptive name for the number of sentence tokens ever produced, then no token of 'L > 0' could have been false. If "making true" is understood in terms of the possible truth of tokens, this will entail that 'L > 0' is deeply necessary, even though it is clearly a posteriori, and will be deeply contingent on the earlier understanding. It seems doubtful that Evans would allow that this sentence is deeply necessary.[27]

Alternatively, we might understand the locution "the sentence would have been true" as invoking an abstract sentence, one that need not be uttered in the state of affairs in question. But we must tread carefully here. It is natural to hold that the abstract sentence 'Julius invented the zip' would have been true in a state of affairs if and only if Julius invented the zip in that state of affairs. But Evans accepts the Kripkean claim that there are states of affairs such that if they obtained, Julius did not invent the zip. If follows from these two claims that the abstract sentence 'Julius

[27] Evans' letter to Martin Davies (this volume) suggests very strongly that he would regard sentences such as 'L > 0' as deeply contingent, and that he would reject a contextual interpretation of deep necessity.

invented the zip' could have been false, so that on this understanding, it is not deeply necessary. So for this strategy to work, Evans must reject the claim that the abstract sentence 'Julius invented the zip' is true in a state of affairs iff Julius invented the zip there, and must give some other account of the evaluation of abstract sentences in states of affairs.[28]

The most natural way to reconcile all this is to interpret the locution "the sentence would have been true" as meaning that the *content* of the abstract sentence would have been true. Then the result above will follow, given Evans' view that the content of "Julius invented the zip" is a descriptive content that differs from the proposition that is contributed to modal contexts. The cost is that on this approach, we cannot use a prior notion of "making true" to *ground* the notion of content, as one can do on some other approaches. Rather, the notion of "making true" and the consequent notions of deep necessity and the like are defined in terms of a prior notion of content.

On Evans' view of content, descriptive names have a descriptive content, because of their epistemic equivalence to descriptions. In this case, "making true" behaves like an epistemic intension. In other cases, it may not. Elsewhere (Evans 1982), Evans rejects the claim that ordinary proper names have a quasi-descriptive content, suggests that the referent of an ordinary proper name is part of its content. On this sort of view, it appears that a sentence such as 'Cicero is Tully' is made true by all states of affairs, so that it is deeply necessary. If this is correct, then deep necessity can come apart from apriority, and Evans' notion of verification does not in general behave in the manner of an epistemic intension.

One source of this difference is that Evans associates content with expression types rather than expression tokens, and imposes a semantic constraint on content: for Evans, the content of an expression type is closely tied to what is required for a speaker of a language to understand it. Given that epistemic intensions are often variable across competently used tokens of an expression type, it follows that epistemic intensions cannot be content in Evans' sense. But in cases such as that of descriptive names, where a specific epistemic intension is required for competent use of an expression type, one can expect that Evans' notion of content will behave more like an epistemic intension, and that deep necessity will coincide more closely with apriority.

Because of all the dependence on a prior notion of content, Evans' account is not naturally assimilated to either an epistemic or contextual understanding of

[28] To support his claims about which sentences could have been true, Evans goes to some length to argue that if *y* had invented the zip and had been F, *y* would have been the referent of 'Julius', and 'Julius is F' would have been true *as a sentence of English*. He argues that there is no *semantic* connection between 'Julius' and a particular referent, so one can suppose that the term could have had a different referent without supposing a semantical change in English. He says: "exactly the same theory of meaning serves to describe the language which would be spoken had *y* invented the zip, as describes the language which is actually spoken" (Evans 1979, 182). These passages use claims about counterfactual *spoken* language to support claims about "making true", so one might initially read them as supporting a contextual interpretation of this notion. But they are arguably also compatible with the second understanding above, if one conjoins this understanding with the thesis that an abstract sentence is true in a state of affairs *if* (although not only if) a token of a semantically identical sentence is true in that state of affairs.

two-dimensionalism. We might think of it as a broadly "semantic" understanding: Evans' first-dimensional modal notions are defined in terms of a prior notion of content, and their grounds depend on the grounds of that notion of content. Given Evans' own distinctive understanding of content, the result is a first-dimensional modal notion that lines up with the epistemic understanding in some cases but not in all. The result is that deep necessity coincides with apriority in some cases (for example, cases involving descriptive names), but not in all cases.

Of course, if one invokes a different notion of content, one will get a different corresponding notion of deep necessity. For example, if one loosens Evans' epistemic constraint on content and embraces a view on which the content of a name just involves its referent, then the corresponding notion of deep necessity may coincide with superficial necessity. And if one loosens Evans' semantic constraint on content and embraces a view on which the content of a token is something like an epistemic intension, then the corresponding notion of deep necessity may coincide with apriority.

5.4 Davies and Humberstone's "fixedly actually"

Davies and Humberstone (1981) give a "formal rendering" of Evans' distinction between deep and superficial necessity, using independently motivated tools from modal logic developed in Crossley and Humberstone (1977). The formal framework starts with a necessity operator N, and supplements it with an "actually" operator **A**, meaning "it is actually the case that". This allows one to represent claims such as 'It is possible for everything which is in fact φ to be ψ', as '$\Diamond\forall x\,(\mathbf{A}\varphi(x) \supset \psi(x))$'. A model theory for these operators requires supplementing the space of possible worlds needed for the necessity operator with a designated "actual world", where $\mathbf{A}\alpha$ is true at a possible world iff α is true at the actual world.

This framework naturally suggests a further idea: just as one can ask whether α is true with respect to a possible world (holding the actual world fixed), one might ask whether α would be true if a *different* world were designated in the actual world. This notion is modeled by adding a further "fixedly" operator **F**, where $\mathbf{F}\alpha$ is true at a world W iff α is true at W no matter which world is designated as actual. Here, the model theory requires that we have a "floating" actual world, or alternatively, it can invoke double-indexed evaluation of sentences at worlds. On the double-indexing approach, we can say that α is true at (V, W) when α is true with respect to W, when V is designated as actual.[29]

[29] What follows is a two-dimensional "reconstruction" of Davies and Humberstone's framework. In their original paper, Davies and Humberstone's official model theory for the system with **F** and **A** involves one-dimensional evaluation with a "floating" actual world, although they note the possibility of a model theory with two-dimensional evaluation.

It is worth observing that it is not obviously correct to say, as is often said, that the ideas of two-dimensional semantics are grounded in two-dimensional modal logic. Two-dimensional modal logic *per se* does not play a crucial role in grounding the frameworks of Kaplan, Stalnaker, and Evans; and even in the case of Davies and Humberstone, two-dimensional modal logic is presented merely as an optional means of representation. Of course many of these ideas can be naturally

The double-indexed evaluation can be formally defined in terms of its interaction with the relevant modal operators.[30] When α is a simple sentence, α is true at (V, W) iff α is true at W according to single-indexed evaluation. When α has the form $\mathbf{F}\beta$, α is true at (V, W) when for all V′, β is true at (V′, W). When α has the form $\mathbf{A}\beta$, α is true at (V, W) when β is true at (V, V). When α has the form $\Box\beta$, α is true at (V, W) when for all W′, β is true at (V, W′). The truth of other complex sentences at (V, W) is the obvious function of the truth of their parts at (V, W). It follows that when α does not contain **F** or **A**, the evaluation of α at (V, W) is independent of V. In these cases, the double-indexed evaluation of α at (V, W) is the same as the single-indexed evaluation of α at W. If α contains **F** or **A**, double-indexed and single-indexed evaluation may come apart.

One can then introduce the combined operator **FA**, which functions so that $\mathbf{FA}\alpha$ is true at (V, W) iff for all worlds V′, α is true at (V′, V′). Or more simply, $\mathbf{FA}\alpha$ is true when α is true at all worlds when that world is designated as actual. Davies and Humberstone note that this operator can be seen as yielding a sort of necessity. We might say that a sentence α is FA-necessary when $\mathbf{FA}\alpha$ is true. For sentences that do not contain **F** or **A**, FA-necessity and ordinary necessity coincide: for all such sentences α, $\mathbf{FA}\alpha$ is equivalent to $\Box\alpha$. But for sentences containing **F** or **A**, FA-necessity and ordinary necessity behave differently.

The **A** operator can be used to represent some contingent a priori truths. For example, if φ is a contingent truth, then $\mathbf{A}\varphi \leftrightarrow \varphi$ is contingent but a priori. Truths of this sort are not necessary, but they are FA-necessary: for example, $\mathbf{FA}(\mathbf{A}\varphi \leftrightarrow \varphi)$ is equivalent to $\Box(\varphi \leftrightarrow \varphi)$, which is true. This behavior parallels Evans' observation that contingent a priori sentences are not superficially necessary, but they are deeply necessary.

Davies and Humberstone extend the parallel between deep necessity and FA-necessity to the case of descriptive names, by suggesting that descriptive names are abbreviations of descriptions of the form 'the actual G', for an appropriate G. On this view, 'Julius' abbreviates 'the actual inventor of the zip'. Then the contingent a priori sentence 'if anyone uniquely invented the zip, Julius did' (which Evans holds is superficially contingent but deeply necessary) is equivalent to 'if anything uniquely has G, the actual G has G'. In Davies and Humberstone's formal framework, this sentence is contingent but FA-necessary. In light of these parallels, Davies and Humberstone put forward the hypothesis that a sentence is deeply necessary in Evans' sense iff it is FA-necessary.

Using this framework, one can define a corresponding sort of 1-intension. We can say the FA-intension of a sentence α is true at W when α is true at (W, W) according to Davies and Humberstone's method of evaluation. In the case of sentences with descriptive names, FA-intensions behave something like a quasi-Fregean semantic

represented using the tools of two-dimensional modal logic, and there is plausibly a relationship between the conceptual bases of these frameworks and of two-dimensional modal logic.

[30] This definition of two-dimensional evaluation is not given explicitly by Davies and Humberstone, but it is easy to see that it gives the intended results.

value. Davies and Humberstone also suggest (without endorsing the suggestion) that one might extend this treatment to other terms, such as 'water', 'red', or 'good', analyzing these as equivalent to 'actually'-involving descriptions.

This raises the question: do FA-intensions satisfy the Core Thesis? Or: is a sentence a priori iff it is FA-necessary? Addressing this sort of question, Davies and Humberstone say that they have found no examples of FA-contingent a priori sentences, and they appear to be sympathetic with the claim that there are no such sentences. But they say that there are many FA-necessary a posteriori sentences, including identities between ordinary proper names ('Cicero = Tully'). This asymmetrical attitude toward the FA-necessary a posteriori and the FA-contingent a priori mirrors Evans' attitude concerning deep necessity. Still, it is prima facie surprising that a formally defined notion such as FA-necessity should yield this asymmetry. So it is worthwhile to assess these claims independently.

A simple approach to these questions runs as follows. Let us say that a sentence of natural language is A-involving if its logical form contains an occurrence of **A** or an occurrence of **F** (in practice there will be few occurrences of **F**, or at least few occurrences of **F** unaccompanied by **A**). If there exist non-A-involving sentences that are necessary a posteriori, then these sentences are FA-necessary a posteriori. If there exist non-A-involving sentences that are contingent a priori, then these sentences are FA-contingent a priori. If no non-A-involving sentences fall into either of these classes, then no A-involving sentences fall into either of these classes. So FA-necessity and apriority are co-extensive if and only if necessity and apriority are coextensive for non-A-involving sentences.

Are there non-A-involving necessary a posteriori sentences? On the face of it, it seems so. For example, identities between ordinary proper names can plausibly be necessary and a posteriori, and such names are plausibly non-A-involving (as Davies and Humberstone themselves note). To resist this claim, one would need to maintain that ordinary proper names, like descriptive names, abbreviate (or are equivalent in logical form to) descriptions of the form 'the actual G'. But there are numerous reasons to doubt such a claim, even on the broadly Fregean view that I have outlined. For example, we have seen that the epistemic intension of a name can vary from speaker to speaker in ways that the epistemic intension of a description does not, so there can be no equivalence in standing meaning between names and descriptions. Even for a single speaker, there may be no expression in the language that encapsulates the epistemic intension of the name as used by that speaker. Further, it is plausible that names have their referents essentially, but descriptions of the form 'the actual G' do not. If this is correct, then ordinary proper names are not A-involving, and identities between them are examples of the FA-necessary a posteriori.

Are there non-A-involving contingent a priori sentences? On the face of it, it seems so. For example, ordinary indexicals such as 'I', 'here', and 'now' give rise to instances of the contingent a priori, and such indexicals are plausibly non-A-involving. For example, 'I am here now (if I exist and am spatiotemporally located)' appears to be both contingent and a priori. Perhaps one could hold that at least one of the indexicals is A-involving: for example, one could suggest that 'I' is equivalent in logical form

to 'the actual speaker', or that 'here' is equivalent in logical form to 'the place where I actually am now'. But these suggestions are unappealing in a number of respects,[31] and are widely rejected in semantics. If these indexicals are not A-involving, then the sentence is plausibly FA-contingent a priori.[32]

These conclusions accord with Davies and Humberstone's view of the FA-necessary a posteriori, but not with their view of the FA-contingent a priori. If correct, these conclusions cast doubt on Davies and Humberstone's claim that FA-necessity is equivalent to Evans' deep necessity. For Evans, the existence of a deeply contingent a priori sentence is "intolerable": it appears to be a conceptual constraint on his notion of content that any a priori sentence has a content that is verified by any state of affairs. If so, and if there are FA-contingent a priori sentences, then it seems that deep necessity is not the same as FA-necessity.

Is the two-dimensional framework of Davies and Humberstone fundamentally a contextual approach or an epistemic approach? As Davies (2004) notes, it seems to be neither. It is clearly not a contextual approach: sentence tokens present in counterfactual worlds play no special role here.[33] And it seems not to be an epistemic approach: epistemic notions play no role in defining the key concepts. I think it is best regarded as a *formal* approach: FA-necessity is in effect defined in terms of its interaction with **A** and **F** operators in a sentence's logical form. This formal definition yields results that are consonant with those of an epistemic interpretation in some cases, but not in all cases.

This consonance stems from the fact that where the **A** and **F** operators are concerned, FA-intensions behave very much like epistemic intensions. If the *only* source of a posteriori necessary and contingent a priori sentences were the **A** and **F** operators, then FA-intensions and epistemic intensions would coincide. But we have seen that these operators appear not to be the only source of these phenomena. Because of this, the two intensions do not coincide, and the Core Thesis fails for FA-necessity.

Of course one might hold that even in the cases of non-A-involving terms that generate a posteriori necessary and contingent a priori sentences, there is something relevant in common with the behavior of A-involving sentences. For example, one might hold that utterances of indexicals and ordinary proper names involve the rigidification

[31] For example, the claim that 'I' is equivalent to 'the actual speaker' has the unappealing consequence that 'if I exist now, I am speaking' is a priori. The 'claim' that 'here' is equivalent in logical form to 'the place where I actually am now' introduces an unappealing asymmetry between the logical forms of 'here' and 'now' that appears to be ad hoc and otherwise unmotivated.

[32] For other examples of non-A-involving contingent a priori sentences, one might try sentences with complex demonstratives or partially descriptive names: for example, 'that picture (if it exists) is a picture', or 'Pine Street (if it exists) is a street'. These sentences are plausibly contingent and are not obviously A-involving. On some views (but not all), these sentences are a priori. If so, these sentences are plausibly FA-contingent a priori.

[33] Surprisingly, Evans seems to have understood Davies and Humberstone's notions as broadly contextual notions. See his letter to Martin Davies (included in this volume), in which he raises "utterance difficulties" for the framework, involving sentences such as 'I exist' and 'There are no speakers'. These parallel the issues raised concerning the contextual understanding in Section 2.4 of this paper.

of some sort of Fregean content, even if these expressions are not equivalent in logical form to corresponding A-involving descriptions. If so, one could use this behavior to define a broader sort of two-dimensional evaluation of sentences that does not turn entirely on the presence of **F** and **A** operators. If one generalized Davies and Humberstone's framework in this way, the resulting framework would more closely resemble the epistemic framework that I have outlined.

5.5 Chalmers' primary intensions

In *The Conscious Mind* (Chalmers (1996), 56–65), I present "a synthesis of ideas suggested by Kripke, Putnam, Kaplan, Stalnaker, Lewis, Evans, and others".[34] I distinguish what the "primary intension" and the "secondary intension" of a concept (where a concept is understood as either a linguistic or a mental token). How are these intensions to be understood? Here I will examine the text from the outside, leaving autobiographical remarks until the end.

The two intensions are initially characterized as follows (p. 57):

> There are two quite distinct patterns of dependence of the referent of a concept on the state of the world. First, there is the dependence by which reference is fixed in the actual world, depending on how the world turns out; if it turns out one way, a concept will pick out one thing, but if it turns out another way, the concept will pick out something else. Second, there is the dependence by which reference in *counterfactual* worlds is determined, given that reference in the actual world is already fixed. Corresponding to each of these dependencies is an intension, which I will call the *primary* and *secondary* intensions, respectively.

The secondary intension seems to be the familiar sort of intension (2-intension, subjunctive intension) across possible worlds. The nature of the primary intension is somewhat less clear. The characterization above has both contextual and epistemic elements. The reference to what "a concept will pick out" under certain circumstances suggests a sort of contextual intension; but reference to how the world "turns out" suggests an epistemic element.

I also say (p. 57) that a concept's primary intension is "a function from worlds to extensions", such that "in a given world, it picks out what the referent of the concept would be if that world turned out to be actual". (The "worlds" are later refined to centered worlds.) This is similar to the characterization above, although the use of "turned out" and "would be" arguably has a slightly different (more subjunctive, less epistemic?) flavor than the use of "turns out" and "will". Again, the referent to potential reference of a concept suggests some sort of contextual intension. This is also suggested by a later discussion (p. 60) which casts the worlds in the domain of a primary intension as Kaplanian "contexts of utterance", and which asks "how things would be if the context of the expression turned out to be W." And again (p. 63):

[34] In retrospect, the "synthesis" remark is unfortunate. As we have seen, the formal similarities between the different frameworks mask deep conceptual differences, which are largely ignored in Chalmers (1996). One moral of this paper is that a blanket citation of theorists who have worked on two-dimensional ideas has the potential to confuse more than it clarifies.

"The primary truth-conditions tell us how the actual world has to be for an utterance to be true in that world; that is, they specify those *contexts* in which the statement would turn out to be true."

If a primary intension is a contextual intension, what sort of contextual intension is it? The discussion suggests an intension that exhibits the sort of quasi-Fregean behavior described in Section 1 of this paper. It seems clear that a primary intension is not intended to be an orthographic contextual intension: nothing in the discussion suggests that in a world where 'water' means steel, the primary intension of our term 'water' picks out steel. It may be intended to be a linguistic contextual intension: footnote 21 on p. 364 suggests sympathy with the view that the word 'water' as used on Twin Earth is of the same linguistic type as ours, in which case a linguistic contextual intension may give quasi-Fregean results. It may also be that some sort of cognitive contextual intension is intended, where one holds fixed the epistemic situation of the subject. But the matter is not clear.

There are also a number of elements in the discussion that suggest an epistemic understanding. The expression "what a concept will refer to if the world turns out" carries an epistemic flavor that is quite different from the subjunctive "what a concept would refer to if the world turned out"; there is arguably more plausibility in the idea that it could *turn out* that 'water' refers to XYZ than that it could have *turned out* that 'water' refers to XYZ. So perhaps there is a sort of amalgam of epistemic and contextual ideas at work in this phrase.

More clearly, the discussion of how to evaluate a primary intension has a strong epistemic element. I say:

> The true intension can be determined only from detailed consideration of specific scenarios: What would we say if the world turned out this way? What would we say if it turned out that way? For example, if it had turned out that the liquid in the lakes was H_2O and the liquid in the oceans was XYZ, then we would probably have said that both were water; if the stuff in oceans and lakes was a mixture of 95 percent A and 5 percent B, we would probably have said that A but not B were water.

Here, the suggestion seems to be that a term's primary intension is constituted to a speaker's or a community's dispositions to apply the term, depending on what is discovered to be the case. This suggests something at least in the vicinity of an epistemic intension. There is still a metalinguistic element in "what would we say?"; for this reason, it seems hard to extend this heuristic to such cases as evaluating "language exists" in a language-free world, and so on. But the idea of capturing the dependence of judgments about extension on discoveries about the actual world suggests something fundamentally epistemic.

Further evidence for an epistemic interpretation stems from two endnotes (notes 26 and 29, p. 366) in which it is stated that one can evaluate a primary intension in worlds that do not contain the original concept. I give the example of "I am in a coma", suggesting that the primary intension should be true of centered worlds where the individual at the center is in a coma and not thinking anything. This is more compatible with an epistemic interpretation than with a contextual interpretation.

The strongest evidence is in footnote 21 (p. 364), which responds to an objector who holds that 'water' on Twin Earth is a different word:

> If one is worried about this . . . one can think of these scenarios as *epistemic* possibilities (in a broad sense) and the conditionals as epistemic conditionals, so that worries about essential properties of words are bypassed.

This response suggests the basis of an epistemic understanding, albeit in quite sketchy terms.[35] This interpretation also fits the claim (p. 64) that a sentence is a priori when it has a necessary primary proposition (where here a proposition is an intension for a statement), and the use of primary intensions to make an inference from conceivability to possibility (roughly, from a claim's a priori coherence to the existence of a world satisfying a claim's primary intension). These moves will be invalid in general if primary intensions are contextual intensions. But if primary intensions are epistemic intensions, it is possible that they are correct.

Autobiographically: I think that primary intensions as I conceived them (both in Chalmers (1995) and in Chalmers (1996)) were much more like epistemic intensions than like contextual intensions. But the distinction between contextual and epistemic understandings was not sufficiently clear in my mind at the time of writing, and is certainly not clear on the page. (The current paper is in part a mea culpa.) Certainly, one must interpret primary intensions as epistemic intensions to make sense of the applications of the two-dimensional framework in these works. (For example, the main conceivability-possibility argument in Chalmers (1996) turns on a version of the thesis of Metaphysical Plenitude outlined earlier.) If one does so, I think the resulting arguments are sound.

5.6 Jackson's A-intensions

Jackson (1998a) discusses "a distinction between two fundamentally different senses in which a term can be thought of as applying in various possible situations". He says:

> We can think of the various situations, particulars, events, or whatever to which a term applies in two different ways, depending on whether we are considering what the term applies to under various hypotheses about which world is the actual world, or whether we are considering what the term applies to under various counterfactual hypotheses. In the first case we are considering, for each world W, what the term applies to in W, given or under the supposition that W is the actual world, our world. We can call this the A-extension of term T in world W—'A' for actual—and call the function assigning to each world the A-extension of

[35] I say a bit more to suggest epistemic evaluation in Chalmers (1995) which develops the framework to yield an account of the narrow content of thought. Here I say that to evaluate a primary intension, one can ask questions such as "If W turns out be actual, what will it turn out that water is"? This sort of "turns-out" conditional is closely related to an indicative conditional, and like an indicative conditional is most naturally interpreted in epistemic terms (see the discussion in Section 3.3 of this paper). Chalmers (1995) suggests that this sort of conditional is superior in a way to "If W is actual, what would the concept refer to?", since it makes clear that the concept is not required to be present in the world. Chalmers (1998) suggests that indicative conditionals can be used to evaluate primary intensions. My current view is that indicative conditionals provide a good heuristic for evaluating primary intensions, but should not be taken as definitional.

T in that world, the A-intension of T. In the second case, we are considering, for each world W, what T applies to in W given whatever world is in fact the actual world, and so we are, for all worlds except the actual world, considering the extension of T in a counterfactual world. We can call this the C-extension of T in W—'C' for counterfactual—and call the function assigning to each world the C-extension of T in that world, the C-intension of T.

Here, the talk of "hypotheses about which world is the actual world", and "given or under the supposition that W is the actual world", strongly suggests that we are thinking about these worlds as a sort of epistemic possibility. One might think for a moment that talk of "what the term applies to under various hypotheses" suggests something contextual, but on reflection there is no more reason why that should be the case here than for the corresponding usage about counterfactual worlds.

Jackson does not say much more about evaluating A-intensions than this. He does say one thing that might suggest a contextual element: he says that the A-proposition (A-intension for a sentence) of 'Some water is H_2O' is contingent, because the sentence is "epistemically possible in the following sense: consistent with what is required to understand it, the sentence might have expressed something both false and discoverable to be false". The claim about what the "sentence might have expressed" strongly suggests something contextual: together with the talk of understanding, it may suggest a sort of cognitive contextual intension. But this locution is not used elsewhere.

One possibility (suggested by conversation with Frank Jackson) is that Jackson's use of the two-dimensional framework rests on a prior commitment to descriptivism. Jackson has argued elsewhere (1998b) that proper names are equivalent to certain rigidified descriptions. If so, then the rigidified description determines a 1-intension (picking out whatever satisfies the unrigidified description in a world) and a 2-intension (picking out whatever actually satisfies the description in all worlds). These intensions will behave just like the name's epistemic and subjunctive intension. The main difference is that on Jackson's approach the framework is used to analyze an independently established aspect of content, whereas on my approach, it is used to independently ground an aspect of content. Viewed this way, Jackson's understanding of the framework might be seen as a semantic understanding akin to Evans', resting on a prior notion of content, although Jackson's conception of the relevant sort of content differs from Evans'.

In any case, most of Jackson's discussion is reasonably consistent with an epistemic understanding of A-intensions. Further, the purposes to which he puts the framework strongly suggest constitutive ties with apriority and the epistemic domains, and epistemic intensions serve these purposes. If so, then A-intensions are arguably identical to epistemic intensions.

5.7 Kripke's epistemic duplicates

Although the work of Kripke (1980) provided the impetus for many of the two-dimensional approaches in the literature, Kripke does not embrace a two-dimensional approach himself. There are numerous remarks that suggest a tacit element of two-dimensional thinking: for example, frequent remarks of the form "*Given* that such-and-such is the case (empirically), such-and-such is necessary." But there is little that formally and explicitly suggests such an approach. While it is possible to analyze many

of Kripke's epistemic claims using possible worlds, Kripke himself generally stays away from this sort of analysis.

There is one exception, however. Kripke notes that in cases where P is the negation of an a posteriori necessary statement, there is some intuition that "it might have turned out that P", even though P is strictly speaking impossible. For example, when a table is made of wood, there is an intuition that the table "might have turned out to be made of ice", even though that is impossible. Kripke denies the modal claim involving "might have turned out" is false in these cases, but he wants to explain it away. Similarly, Kripke notes that for statements such as 'heat is the motion of molecules', there is a sense of "apparent contingency", even though the statement is strictly necessary. Again, Kripke wants to explain this sense of contingency away.

Kripke suggests the following strategy. In these cases, although a statement is necessary, we can say that *under appropriate qualitatively identical evidential situations, an appropriate qualitatively identical statement might have been false*. And he suggests this explains the sense of apparent contingency.

> What, then, does the intuition that the table might have turned out to be made of ice ... amount to? I think it means simply that there might have been *a table* looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. (1980, 142)

He applies a similar strategy to 'heat is the motion of molecules', and other cases. The general principle is that when there is an intuition of apparent contingency associated with a necessary truth P, there is a qualitatively identical contingent truth P*, such that P* might have been false in an evidential situation qualitatively identical to the original situation. Where P is 'heat is the motion of molecules', P* might be 'heat sensations are caused by the motion of molecules'.

This apparently innocuous principle packs considerable power: it enables us to reason from epistemic premises to modal conclusions. When P is "apparently contingent", or such that it seems that "it might have turned out that P", P has a distinctive *epistemic* status: to a first approximation, these claims come to the claim that P is not ruled out a priori. But the conclusion here is a *modal* one: that a certain state of affairs (involving a subject, evidence, and a statement) might really have obtained. In effect, Kripke reasons from a premise about the epistemic status of a statement to a conclusion about the possible truth of a statement token that shares a type with the original statement.

This reasoning can be modeled using the two-dimensional framework, understood contextually. We might say that the *evidential contextual intension* of a given statement is a function that is defined at centered worlds in which there is a subject with qualitatively identical evidence, uttering a qualitatively identical statement, and that returns the truth-value of that statement. Then Kripke is in effect suggesting that when a statement is "apparently contingent", its evidential contextual intension is contingent.

The intension in question is not fully defined, as Kripke does not define what it is for evidential situations or statements to be qualitatively identical. But it is natural to suggest that two evidential situations are identical when they are phenomenologically

equivalent: that is, when what it is like to be in the first situation is the same as what it is like to be in the second. As for qualitatively identical statements: one might first suggest that this occurs when they have similar descriptive or Fregean content, but that suggestion might be inappropriate in the current context. An alternative suggestion is that two statements are qualitatively identical when they (or corresponding thoughts) play similar cognitive roles for the subject. This goes beyond Kripke and is somewhat loose, but it seems at least compatible with his discussion.

Reconstructed this way, Kripke's principle takes on a familiar shape. In effect, the claim is that when a statement is apparently contingent, it has a contingent evidential contextual intension. If one substitutes aposteriority for apparent contingency and rearranges a little, one gets a familiar-looking result: if a statement has a necessary evidential contextual intension, it is a priori.

So Kripke's principle here suggests something in the vicinity of the Core Thesis for an evidential contextual intension. This should raise alarm bells. We have already seen that the Core Thesis appears to be false for any sort of contextual intension. Applying the sort of reasoning from our earlier discussion, one can straightforwardly come up with a counterexample. 'I have such-and-such evidence' is one example. For a more interesting example, let 'Bill' be a name that rigidly designates the phenomenological quality instantiated at the center of my visual field. Let us say that that quality (for me now) is phenomenal blueness. Then 'Bill is phenomenal blueness' is plausibly a posteriori, but it has a necessary evidential contextual intension.

This point is not simply an artifact of our reconstruction; it applies equally to Kripke's original claim. 'Bill is phenomenal blueness' is apparently contingent in the same sort of way as paradigmatic apparently contingent statements. We have an intuition that it might have turned out that Bill was not phenomenal blueness—it might have been that Bill was phenomenal redness, for example. This intuition seems to be on a par with our intuitions about the table, heat, and so on. But there is no qualitatively identical evidential situation in which a qualitatively identical statement would have been false. So the Kripkean reasoning is invalid in this case.

Perhaps Kripke could shrug this off and say that the reasoning was never intended to apply in *all* cases. But this reply would have a significant cost, since the reasoning is crucial to Kripke's argument against mind–body identity theories. We have a strong intuition of apparent contingency associated with 'pain is C-fibers'; and Kripke argues that this cannot be explained away by the claim that in a qualitatively identical evidential situation, a qualitatively identical statement would have been false:

> To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of pain *is* not to have a pain. The apparent contingency of the connection between the physical state and the corresponding brain state thus cannot be explained by some sort of qualitative analog as in the case of heat. (1980, 152)

Kripke uses this point to argue that as the apparent contingency cannot be explained away, the claim that 'pain is C-fibers' is not necessary at all. It follows that the claim is not true, since if it is true, it must be necessary. But it is clear that everything in the quoted passage applies equally to 'Bill'. To be in the same epistemic situation that

would obtain if one had Bill *is* to have Bill; and so on. So the 'Bill' case equally cannot be explained by Kripke's paradigm. But here, one *cannot* reason from failure to satisfy Kripke's paradigm to contingency and thus to falsity: 'Bill is phenomenal blueness' is clearly true and necessary. So by parity, one cannot apply this reasoning in the case of 'pain'. Insofar as Kripke's argument against the mind–body identity theory relies on this reason, it appears that the argument is unsound.

The moral here seems to be that there is nothing special about one's *evidence* in diagnosing apparent contingency. This suggests a natural response: one might *weaken* Kripke's principle, holding simply that when a statement is apparently contingent, an appropriate qualitatively identical statement might have been false. This claim would cover the standard Kripkean cases, and would also encompass cases involving contingency of evidence itself, such as that of 'Bill'. One might think that to do this would be to kill off the argument against the identity theory, which relied crucially on qualitatively identical evidence. But there is another strand in Kripke's argument that might still apply.

Immediately after noting the considerations above, Kripke says:

> The same point can be made in terms of the notion of what picks out the reference of a rigid designator. In the case of the identity of heat with molecular motion the important consideration was that although 'heat' is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us sensation S. It is thus possible that a phenomenon should have been rigidly designated in the same way as a phenomenon of heat, with its reference also picked out by means of the sensation S, without that phenomenon being heat and therefore without its being molecular motion. Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality. Thus pain, unlike heat, is not only rigidly designated by 'pain' but the reference of the designator is determined by an essential property of the referent.

Kripke says that this is the "same point", but in fact it is not. A close examination suggests that this point has nothing to do with evidence: it does not rely on the notion of a statement in a qualitatively identical evidential situation, but rather on that of a designator with the same manner of reference-determination. So we can apply the weakened principle suggested above, where a "qualitatively identical statement" is understood, perhaps, as a statement with the same reference-fixing intensions. Under this paradigm, the apparent contingency of 'Bill is phenomenal blue' *can* be explained away, as it is possible for a statement associated with the same reference-fixing intentions—to refer to the quality in the center of one's visual field—to be false, for example, in a case where the quality is phenomenal red. But the case of 'pain' *cannot* be explained away in these terms. So the argument against the identity theory would now seem to go through.

Nevertheless, more problems immediately arise. The reasoning here in effect invokes the claim that an apparently contingent statement has a contingent *qualitative contextual intension*, where this sort of intension is defined at worlds centered on a statement qualitatively identical to the original, returning the referent of that token. For familiar reasons, there are counterexamples to this claim. One such is 'A sentence token exists'. Or let 'L' be a descriptive name that rigidly designates

the number of actual (spoken or written) languages. Then 'L > 0' is apparently contingent: it seems that it might have turned out that L was zero, and so on. But 'L > 0' has a necessary fixing contextual intension. So the principle is false. Further, it is easy to see that even the weakened Kripkean reasoning suggested above does not apply in the case of 'L > 0': it is not possible for a qualitatively identical statement to be true, but the statement is necessary all the same. So even the weakened reasoning, from failure to satisfy the weaker paradigm to contingency, is invalid. So although Kripke's second argument against the identity theory is different from the first argument, it is also invalid.[36]

I think that the moral in both cases is that Kripke's diagnosis of intuitions about apparent contingency is incorrect: they do not turn essentially on intuitions about qualitatively identical evidential statements, and they do not turn essentially on intuitions about qualitatively identical statements.[37] Rather, they turn on intuitions about the direct evaluation of epistemic possibilities. And we have seen that this sort of epistemic evaluation does not turn essentially on the status of possible tokens. This suggests in turn that to reason about apparent contingency, then instead of the Kripkean principles above, which tacitly involve a contextual understanding, one should appeal to principles involving an *epistemic* understanding: for example, that when a statement is apparently contingent, there is some scenario (considered as actual) in which it is false. Or, if we understand scenarios as centered worlds: that when a statement is apparently contingent, there is some centered world in which the statement's epistemic intension is true. I think that once we understand things this way, an argument against the mind–body identity theory of a sort analogous to Kripke's has a chance of succeeding. But I have written about that elsewhere.[38]

In any case, we can see that while Kripke does not explicitly endorse a two-dimensional approach, these issues are quite close to the surface in his discussion. One can even see here an instance of a familiar two-dimensionalist pattern: trying to capture epistemic phenomena with a contextual approach, coming close, but not

[36] Bealer (1996) develops Kripke's proposal using his notion of a "semantically stable" expression (one such that necessarily, in any language group in an epistemic situation qualitatively identical to ours, the expression would mean the same thing). Bealer suggests that semantically stable expressions are invulnerable to "scientific essentialism". It is not clear exactly what is required for a "qualitatively identical epistemic situation" for a language group, but it is clear that however this notion is understood, terms such as 'L' above will be semantically stable, as will rigid designators for aspects of a group's epistemic situation. So there will be a posteriori necessities involving semantically stable expressions. As in Kripke's discussion and in other cases, a broadly contextual notion (semantic stability) serves as an imperfect substitute for a broadly epistemic notion (semantic neutrality).

[37] There is arguably a strand in the passage above that does not turn directly on qualitatively identical statements. This is the claim that while heat is picked out by a property it has accidentally, pain is picked out by a property it has essentially. This suggests that Kripke's "same point" may come down to three different points, with potentially three different arguments in the background. I think that the third argument has the most chance of success. But I think that to make the argument work in the general case, once has to adopt something like the (non-Kripkean) framework of epistemic intensions.

[38] See, e.g., Chalmers (2002a, 2003).

quite succeeding. I think that once again, the moral is that the epistemic framework is most fundamental for these purposes.

5.8 Other approaches

There are a number of other approaches that either fit within a two-dimensional framework or have something in common with these ideas. I will discuss some of these briefly.

Tichy (1984) suggests that there are two propositions corresponding to any given statement: the proposition expressed and the proposition associated. The proposition expressed by 'Phosphorus is hot' is the proposition that Venus is hot. This proposition behaves like a 2-intension. The proposition associated with 'Phosphorus is hot' is the proposition that the sentence 'Phosphorus is hot' says in English that Phosphorus is hot. This proposition behaves like a (modified) linguistic contextual intension: it is true at those worlds corresponding to a centered world where the sentence's linguistic contextual intension is true, and false at all other worlds.

Tichy uses this distinction to argue that there are no necessary a posteriori propositions. He suggests that for truths such as 'Hesperus is Phosphorus', the (trivial) proposition expressed will be necessary and a priori, while the proposition associated will be contingent and a posteriori. Whether the latter is so depends on how sentences and languages are individuated: if 'Hesperus' picks out Venus essentially, then the proposition associated will be necessary. Even if this possibility is set aside (as Tichy does), there will clearly be cases (e.g. 'A sentence token exists') of an intuitively necessary a posteriori statement with a necessary associated proposition.[39] Tichy nowhere says that *all* intuitively a posteriori statements have an a posteriori associated proposition, and his claim that there are no necessary a posteriori propositions is arguably independent of this claim.[40] But this suggests at least that Tichy's dual-proposition account of the Kripkean phenomena does not quite get to the roots of the phenomena.

Also relevant are some proposals for assigning content to *thought* rather than to language.[41] One such proposal is Lewis's (1979; 1986; 1994) suggestion that a subject's system of belief can be modeled by the self-attribution of a property, or equivalently,

[39] Tichy sets this possibility aside on the (highly arguable) grounds that Kripke is committed to denying that it is necessary that 'Hesperus' in English picks out Venus, since on Kripke's view all that is essential to 'Hesperus' in English are linguistic conventions involving reference-fixing.

[40] The claim that there are no necessary a posteriori propositions is sometimes regarded (e.g. Byrne 1999; Soames 2004) as a central aspect of the two-dimensional framework. I think it should be regarded as a further inessential claim. If one stipulates that propositions are sets of possible worlds, such a claim may be reasonable. On my own view, it is better to regard propositions as the contents of sentences (leaving open their nature) and allow that the apriority and the necessity of a statement correspond respectively to the apriority and the necessity of the proposition it expresses. If so, some propositions will be both necessary and a posteriori, and propositions will presumably themselves have a sort of two-dimensional modal structure (although they may not be reducible to this structure).

[41] Apart from those mentioned in the text, some other relevant proposals concerning the narrow content of thought include the "immediate object of belief" of Brown (1986) (roughly: those propositions believed by all intrinsic duplicates), the "notional worlds" of Dennett (1982) (roughly:

by a class of possible individuals (the subject's "doxastic alternatives"), or by a class of centered worlds. Lewis (1979) assigns this sort of content to specific beliefs, while Lewis (1986; 1994) suggests a content of this sort can be used to represent a subject's total belief system, and then in turn to show how the subject satisfies various belief ascriptions. It is clear from Lewis's discussion that this set of centered worlds behaves at least something like a 1-intension: a subject who believes *I am hungry* will have only worlds centered on hungry subjects in the set; a subject who believes *water is XYZ* will have Twin Earth centered worlds in the set, and so on.

Lewis's proposal is hard to classify directly in the present system, since he does not say much about how the relevant set of worlds (or the relevant property, or the relevant class of alternatives) is defined, apart from saying that it is determined by the subject's behavior and functional organization by the principles of belief-desire psychology (Lewis 1986, 36–40).[42] Lewis says nothing to suggest that a token of specific mental or linguistic states is required at the center of the relevant worlds, however, so there is reason to think he is not invoking a contextual understanding. And Lewis's centered worlds can naturally be seen as representing a sort of epistemic possibility for the subject. So Lewis's discussion seems at least consistent with an epistemic understanding, on which the relevant set of centered worlds is a sort of epistemic intension of a subject's total belief state, consisting to a first approximation of those scenarios that verify all of a subject's beliefs.

White (1982) sets out a multi-dimensional proposal for understanding the narrow (internally determined) content of thought and language. To simplify a little, White defines the *partial* character of a word as a function from "contexts of acquisition" (worlds centered on a functional duplicate of the original subject) to the Kaplanian character of the corresponding word in that context. This is actually a three-dimensional function (a function from centered worlds to contexts to worlds to extensions), but one can diagonalize it twice to yield a one-dimensional function: in effect, a functional contextual intension, defined at worlds centered on a functional duplicate of the original subject, returning the extension of the relevant token.

The resulting functional contextual intension will be internally determined by definition, and will be a reasonably good approximation to a Fregean semantic value. It will give anomalous results in a few cases: for example, cognitively significant claims concerning language ('A sentence token exists') or a subject's functional organization ('I have computational structure C') may have a necessary functional contextual intension, and so will have a relatively trivial partial character. But if we assume that the epistemic intensions of a subject's tokens are determined by the

the set of environments in which the organism as currently constituted will flourish), the "realization conditions" of Loar (1988) (roughly: the set of worlds in which a given thought would be true if it were not a misconception), and the "notional content" of White (1991) (roughly: the class of worlds for which a subject's actions are optimal). All of these have some similarity to 1-intensions, characterized in a broadly contextual way.

[42] For example, the principle that a subject's actions are such that if the subject's beliefs were true, they would tend to fulfill the subject's desires. One can see this principle as mutually constraining a subject's classes of doxastic alternatives and desire alternatives.

subject's functional state, then the diagonalized partial character will give a reasonable approximation of an epistemic intension.

Fodor (1987) gives a related proposal for understanding narrow content. He suggests that the narrow content of a thought is a function from contexts to truth-conditions, where contexts appear to behave like worlds centered on the thought, and truth-conditions behave like 2-intensions. Like White's partial character, Fodor's function is not directly truth-evaluable (this led Fodor to eventually reject the proposal on the grounds that it is not really a sort of content), but as usual one can diagonalize it to yield a truth-evaluable content. The result is something like a conceptual contextual intension, mapping worlds centered on a token of the relevant concept or thought to that token's extension.

As with conceptual (and linguistic) contextual intensions in general, the behavior of Fodor's function will depend on how concepts and thoughts are individuated, in order to know which centered worlds are relevant. If they are syntactic mental types, then one has a sort of orthographic intension, which has uninteresting content. If they are semantic types, then it is unclear how one can specify the relevant semantic value in a noncircular way. These points (and many others) are developed by Block (1991), in a thorough critique of proposals of this sort in accounting for narrow content.

Block makes a point worth noting here: he suggests that proposals involving a mapping from contexts "often seem to engender a *cognitive illusion* to the effect that we know *what the proposed mapping is*". I diagnose things differently. Our judgments about the mapping are grounded in perfectly reasonable intuitions about the evaluation of our terms in epistemic possibilities. The illusion on the part of these theorists is not that they grasp the mapping. Rather the illusion is that the mapping is grounded in context-dependence. Once we recognize the epistemic roots of the mapping, the problems go away.

Whether or not this diagnosis is correct, it seems fair to say that in many of the cases we have discussed, various contextual two-dimensional notions are of interest largely to the extent that they approximate epistemic two-dimensional notions. One might regard contextual notions positively as an inexpensive substitute for the epistemic notions, yielding many of the benefits without as many of the costs. Or one might regard them negatively, as a "distractor" from the more important epistemic notions that can lead to confusion because of their surface similarity. I suspect that there is something to each of these attitudes. But in any case, it seems clear that the epistemic understanding yields semantic notions with the deepest connections to the cognitive, rational, and epistemic domains.

6. Conclusion

One might see the project of this paper as an attempt to vindicate Carnap's vindication of Frege, although a number of aspects of the approach diverge from both. I have followed Carnap's lead in using a modal analysis to construct a semantic value that is constitutively tied to the epistemic domain. If the project is successful, it yields an aspect of meaning that can serve as the third vertex in the golden triangle, by virtue of its constitutive connections to reason and modality.

I have taken epistemic notions as primitive in this paper, and have constructed semantic notions from there. But it should be stressed that proceeding in this way does not entail that one of the vertices of the golden triangle is more fundamental than the others. I have appealed to this order of explication because we have a relatively pretheoretical grasp on the relevant epistemic notions, whereas we do not have the same pretheoretical grasp on the corresponding semantic notions. But this no more entails that reason is prior to meaning than our pretheoretical grasp of the macroscopic world entails that it is prior to the microscopic world.

The framework in this paper is compatible with a variety of views about the underlying relations between epistemic and semantic notions. It could be that epistemic properties are grounded in semantic properties (so that thoughts, for example, stand in epistemic relations by virtue of their epistemic content), or it could be that semantic properties are grounded in epistemic properties (so that thoughts have their semantic content in virtue of their epistemic role). Or it could be that, as I am inclined to suspect, neither is more basic than the other. In any case, we can expect that two-dimensional semantics will be helpful in analyzing the complex connections between meaning, reason, and modality.

References

Bealer, G. (1996). A priori knowledge and the scope of philosophy. *Philosophical Studies* 81: 121–42.

Block, N. (1991). What narrow content is not. In B. Loewer and G. Rey, eds., *Meaning in Mind: Fodor and his Critics*. Blackwell.

——and Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108: 1–46.

Braun, D. (1996). Demonstratives and their linguistic meanings. *Noûs* 30: 145–73.

Brown, C. (1986). What is a belief state? *Midwest Studies in Philosophy* 10: 357–78.

Byrne, A. (1999). Cosmic hermeneutics. *Philosophical Perspectives* 13: 347–83.

——and Pryor, J. (2006). Bad intensions. In M. Garcia-Carpintero and J. Macia, eds, *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity*. University of Chicago Press.

——(1963). Replies and systematic expositions. In P. Schilpp, ed., *The Philosophy of Rudolf Carnap*. Open Court.

Chalmers, D. J. (1995). The components of content (1995 version). Manuscript (revised as Chalmers 2002c). [consc.net/papers/content95.html]

——(1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

——(1998). The tyranny of the subjunctive. [consc.net/papers/tyranny.html]

——(1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research*. [consc.net/papers/modality.html]

——(2002a). Does conceivability entail possibility? In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility*. Oxford University Press. [consc.net/papers/conceivability.html]

——(2002b). On sense and intension. [consc.net/papers/intension.html]

——(2002c). The components of content (revised version). In *Philosophy of Mind: Classical and Contemporary Readings*. [consc.net/papers/content.html]

——(2003). Consciousness and its place in nature. In S. Stich and F. Warfield, eds., *Blackwell Guide to Philosophy of Mind*. Blackwell. [consc.net/papers/nature.html]
——(forthcoming). The nature of epistemic space. [consc.net/papers/espace.html]
——and Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review* 110: 315–61. [consc.net/papers/analysis.html]
Crossley, J. N. and Humberstone, I. L. (1977). The logic of 'actually'. *Reports on Mathematical Logic* 8: 11–29.
Davies, M. and Humberstone, I. L. (1981). Two notions of necessity. *Philosophical Studies* 58: 1–30.
——(2004). Reference, contingency, and the two-dimensional framework. *Philosophical Studies* 118: 83–131.
Dennett, D. C. (1982). Beyond belief. In A. Woodfield, ed., *Thought and Object*. Oxford University Press.
Evans, G. (1979). Reference and contingency. *The Monist* 62: 161–89.
——(1982). *The Varieties of Reference*. Oxford University Press.
Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
Frege, G. (1892). Über Sinn und Bedeutung. Translated in P. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1952.
Garcia-Carpintero, M. (1998). Indexicals as token-reflexives. *Mind* 427: 529–63.
Jackson, F. (1998a). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.
——(1998b). Reference and description revisited. *Philosophical Perspectives* 12: 201–18.
Kaplan, D. (1978). Dthat. In P. Cole, ed., *Syntax and Semantics*. New York: Academic Press.
——(1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. Oxford: Oxford University Press.
Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
Lewis, C. I. (1944). The modes of meaning. *Philosophy and Phenomenological Research* 4: 236–50.
Lewis, D. (1979). Attitudes *de dicto* and *de se*. *Philosophical Review* 88: 513–43.
——(1986). *On the Plurality of Worlds*. Blackwell.
——(1994). Reduction of mind. In S. Guttenplan, ed., *Companion to the Philosophy of Mind*. Oxford: Blackwell.
Loar, B. (1988). Social content and psychological content. In R. H. Grimm and D. D. Merrill, eds., *Contents of Thought*. University of Arizona Press.
Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson, ed., *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.
Salmon, N. (1986). *Frege's Puzzle*. MIT Press.
Schiffer, S. (1990). The mode-of-presentation problem. In C. A. Anderson and J. Owens, eds., *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*. Stanford: CSLI Press.
Schroeter, L. (2004). The rationalist foundations of Chalmers' two-dimensional semantics. *Philosophical Studies*, 118: 227–55.
Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press.
——(2004). *Reference and Description*. Princeton University Press.
Stalnaker, R. (1978). Assertion. In P. Cole, ed., *Syntax and Semantics: Pragmatics, Vol. 9*. New York: Academic Press.

Stalnaker, R. (1990). Narrow content. In C. A. Anderson and J. Owens, eds., *Propositional Attitudes*. Stanford: Center for the Study of Language and Information.

——(1991). How to do semantics for the language of thought. In B. Loewer and G. Rey, eds., *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell.

——(1999). *Content and Context* (foreword). Oxford University Press.

——(2001). On considering a possible world as actual. *Aristotelian Society Supplementary Volume* 75: 141–56.

Tichy, P. (1984). Kripke on necessity a posteriori. *Philosophical Studies* 43: 225–41.

Weatherson, B. (2001). Indicatives and subjunctives. *Philosophical Quarterly* 51: 200–16.

White, S. (1982). Partial character and the language of thought. *Pacific Philosophical Quarterly* 63: 347–65.

——(1991). Narrow content and narrow interpretation. In *The Unity of the Self*. MIT Press.

Yablo, S. (1999). Concepts and consciousness. *Philosophy and Phenomenological Research* 59: 455–63.

——(2002). Coulda, woulda, shoulda. In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility*. Oxford University Press.

5

Reference, Contingency, and the Two-Dimensional Framework

Martin Davies

Near the beginning of 'Reference and contingency', Gareth Evans says (1979: 178):

> This paper is an attempt to [use] a puzzle about the contingent *a priori* to test and explore certain theories of reference and modality. No one could claim that the puzzle is of any great philosophical importance by itself, but to understand it, one has to get clear about certain aspects of the theory of reference; and to solve it, one has to think a little more deeply than one is perhaps accustomed about what it means to say that a statement is contingent or necessary.

The most familiar examples of the puzzle of the contingent *a priori* and the mirror-image puzzle of the necessary *a posteriori* involve what appear to be referring expressions: ordinary proper names, names of natural kinds, names with their reference fixed by description. So an account of the puzzles can scarcely avoid involvement with the theory of reference. But, as Evans stresses, there are other examples of the contingent *a priori* and the necessary *a posteriori* that do not involve referring expressions at all. So no thesis about reference can suffice, by itself, for a complete solution to the puzzles. Rather, Evans proposes, a solution must be provided by reflection on the modal notions of contingency and necessity.

Evans's response to the puzzle about the contingent *a priori* makes use of a distinction between 'superficial' and 'deep' notions of necessity. In 'Two notions of necessity', Lloyd Humberstone and I suggested that Evans's distinction could be rendered by a distinction between two operators in two-dimensional modal logic.[1] The present paper is, on the one hand, a review and reconsideration of some of the themes of 'Two notions of necessity' and, on the other hand, an attempt to reach a deeper understanding and appreciation of Evans's reflections on both modality and reference. The

Thanks to David Braddon-Mitchell, David Chalmers, Manuel García-Carpintero, Frank Jackson, Fred Kroon, and Daniel Stoljar for comments and conversations. My greatest and longest-standing debts are, of course, to Lloyd Humberstone and Gareth Evans.

[1] Davies and Humberstone (1980). In fact, the model-theoretic semantics presented in that paper is not explicitly two-dimensional. It makes use, instead, of a notion of *variance* between (one-dimensional) models for a modal language with an 'Actually'-operator. In such models, there is a distinguished world, w*, and two models stand in the relation of variance if they differ at most over which world is the distinguished world. The equivalence between this way of presenting the model theory and the two-dimensional way is noted at p. 26, n. 4.

aim, in very general terms, is to plot the relationships between the notions of necessity that Humberstone and I characterized in terms of two-dimensional modal logic, the notions of necessity that Evans himself described, and the epistemic notion of *a priority*. I begin with the two-dimensional framework as Humberstone and I conceived of it.

1. The Two-Dimensional Framework

It is a familiar point that there are natural-language sentences, such as 'It is possible that everything that is actually red should have been shiny', that resist formulation given just the standard resources of a quantified modal language.[2] In the case of this example, one of the two obvious candidates:

$$\Diamond(\forall x)(x \text{ is red} \rightarrow x \text{ is shiny})$$

is inadequate because it requires, as the original sentence does not require, that in the envisaged possibility things that are red should *also* be shiny. The other obvious candidate:

$$(\forall x)(x \text{ is red} \rightarrow \Diamond(x \text{ is shiny}))$$

is also inadequate because it fails to require, as the original sentence does require, that in the envisaged possibility the things that are actually red should be shiny *together*.

1.1 Introducing 'Actually'

It is an equally familiar point that the solution to this expressive inadequacy is to introduce an 'Actually'-operator, 'A'. In terms of possible-worlds model-theoretic semantics for the modal language, a sentence 'As'[3] is true with respect to a possible world, w, just in case the embedded sentence s is true with respect to the model's designated or 'actual' world, w*. In terms of homophonic truth-conditional semantics, 'As' is true just in case s is actually true. With the help of this new operator, the originally problematic natural-language sentence can be formalized as:

$$\Diamond(\forall x)(\mathsf{A}(x \text{ is red}) \rightarrow x \text{ is shiny}).$$

This sentence is true in a model just in case there is some possible world, w, such that each object that is red with respect to the model's actual world, w*, is shiny with respect to w. All the objects that are red in w* are required to be shiny *together*—that

[2] I first learned about the logic of 'actually' from Lloyd Humberstone in 1974. See Crossley and Humberstone (1977), from which this example and much else is borrowed. See also Hazen (1976), and Humberstone (2004), section 2.

[3] Where confusion is unlikely to result, I use ordinary quotation marks even though corner quotes would be more accurate. What is intended here is 'A^s', the concatenation of the 'Actually'-operator, 'A', with the sentence s.

is, in the same possible world, w—but nothing is required to be red in w and *also* shiny in w.[4]

The semantic rule for the 'Actually'-operator, 'A', has the result that if 'As' is true with respect to any world then it is true with respect to every world. So if 'As' is true then so is '$\Box(\mathsf{A}$s)'. While this is an immediate consequence of the intuitive semantics for 'A', it does not accord well with the idea that it is a largely contingent matter what is actually the case. Suppose, for example, that the embedded sentence s means that the earth moves, and that this is contingently true. Then, even allowing that there is a notion of necessity expressed by the modal operator '$\Box$' on which 'As' is necessarily true (that is, '$\Box(\mathsf{A}$s)' is true), we also want to say that there is another notion of necessity on which 'As' is not necessarily true. This second notion of necessity is needed to capture the intuition that it is a contingent matter which possible world is actual.

1.2 Introducing 'Fixedly'

In response to this intuition about a second notion of necessity, Davies and Humberstone (1980) proposed that a further operator '$\mathcal{F}$' ('Fixedly') be added to modal languages, alongside both '$\Box$' and 'A'.[5] Thus, while the introduction of the 'Actually'-operator is motivated by issues about expressive inadequacy, this is not so for the introduction of the 'Fixedly'-operator.

Just as '$\Box$' universally quantifies over possible worlds playing the role of the world with respect to which truth is being evaluated, '$\mathcal{F}$' universally quantifies over worlds playing the role of the actual world—the world to which the operator 'A' directs us. So now we allow for variation both in the world of evaluation, w_j, and in the world playing the role of the actual world, w_i. A sentence '$\Box$s' is true with respect to a world w_j with world w_i playing the role of the actual world just in case, for every world w, the embedded sentence s is true with respect to w, with world w_i still playing the role of the actual world. A sentence '$\mathcal{F}$s' is true with respect to world w_j with world w_i playing the role of the actual world just in case, for every world w, the embedded sentence s is true with respect to w_j, but now with w playing the role of the actual world. The operator '$\mathcal{F}$' by itself does not capture an intuitive notion of necessity at all for, if the embedded sentence s contains no occurrences of 'A', then '$\mathcal{F}$s' is simply equivalent to s. But once '$\mathcal{F}$' is available we can explore the properties of the combination '$\mathcal{F}\mathsf{A}$'.

In this framework, the semantic rule for 'A' says that 'As' is true with respect to w_j with w_i playing the role of the actual world just in case the embedded sentence s is true with respect to w_i, with w_i still playing the role of the actual world. If we put this together with the semantic rule for '$\mathcal{F}$', the result is that '$\mathcal{F}\mathsf{A}$s' is true with respect to w_j with w_i playing the role of the actual world just in case, for every world w, the embedded sentence s is true with respect to w, with w *also* playing the role

[4] Here, I set aside two complications. One concerns the interpretation of first-order quantification when the domain of objects varies from possible world to possible world. The other concerns the use of second-order quantification in order to overcome the expressive limitations of quantified modal languages without an 'Actually'-operator. See Forbes (1989), esp. chs 2 and 4.

[5] Davies and Humberstone (1980), n. 7 and Bibliography item [11], referred to a paper, 'The logic of "Fixedly" ', as forthcoming. In fact, this material appeared as Appendix 10 of Davies (1981).

of the actual world. So the initial pair of worlds $<w_i, w_j>$ does not matter. A sentence '$\mathcal{F}$A's' is true just in case, for every world w, s is true with respect to w, with w also playing the role of the actual world. We might express this by saying that '$\mathcal{F}$A's' is true just in case, for every world w, s is true at w *considered as actual* (Davies and Humberstone 1980, p. 3).

If the embedded sentence s contains no occurrences of 'A', then '$\mathcal{F}$A's' is equivalent to '□s'. But '$\mathcal{F}$A' is not in general equivalent to '□', as we can see if we consider its application to the problematically necessary sentence 'As'. While '□(As)' is true, '$\mathcal{F}$A(As)' is equivalent to '$\mathcal{F}$As' and so to '□s', which is false. Davies and Humberstone thus proposed that, in a modal language with '$\mathcal{F}$' alongside the more familiar '□'and 'A', we could express two notions of necessity. There is one notion, expressed by '□', for which 'As' is, if true, then necessarily true; and there is another notion, expressed by '$\mathcal{F}$A', for which 'As' is, though true, not necessarily true (unless s is itself necessarily true). This second notion thus captures the intuition that it is a contingent matter which possible world is actual.

1.3 Two-dimensional arrays and one-dimensional intensions

It is clear why modal logic with 'A' and '$\mathcal{F}$' is *two-dimensional* modal logic. In any model, the evaluation function for a sentence is a mapping from pairs of possible worlds to truth values (a *2D-intension*). In each pair, one world plays the role of the actual world and one plays the role of the 'floating' world. The evaluation of a sentence can thus be represented in a two-dimensional array, in which each row is labelled with a world playing the role of the actual world and each column is labelled with a world playing the role of the floating world.

Before '$\mathcal{F}$' was introduced we only needed to consider the top row of such an array, for the world playing the role of the actual world was held constant (as $w^* = w_1$, say). Now, with '$\mathcal{F}$' added to the language, the truth of '□s' with respect to the pair of w_i as the actual world and w_j as the floating world requires that s be evaluated as true in each cell on the w_i-labelled row. The truth of '$\mathcal{F}$s' with respect to the same pair requires that s be evaluated as true in each cell on the w_j-labelled column. The truth of '$\mathcal{F}$□s' with respect to any pair requires that s be evaluated as true in every cell in the two-dimensional array. And finally, the truth of '$\mathcal{F}$As' with respect to any pair requires that s be evaluated as true in each cell on the leading diagonal. Thus, in the two-dimensional framework, the necessity expressed by '$\mathcal{F}$A' is *truth on the diagonal*.

It is natural to associate with each sentence, in addition to its 2D-intension, three one-dimensional intensions or mappings from possible worlds to truth values. First, corresponding to the w_i-labelled row there is the horizontal intension (*H-intension*) for w_i as the actual world. The H-intension for the original actual world, $w^* = w_1$, might be called the H-intension *simpliciter*. Second, similarly, corresponding to the w_j-labelled column there is the vertical intension (*V-intension*) for w_j as the floating world; and the V-intension for w_1 is the V-intension *simpliciter*. Third, corresponding to the diagonal, there is the *D-intension*. Of these three one-dimensional intensions, it is the H-intension and the D-intension that will mainly concern us in what follows. The H-intension corresponds to Chalmers's secondary intension, Jackson's C-intension, and Stalnaker's 'what is said'; the D-intension

to Chalmers's primary intension, Jackson's A-intension, and Stalnaker's diagonal proposition (Chalmers, 1996; Jackson, 1998a; Stalnaker, 1978).

Both H-intensions and D-intensions are functions from worlds to truth values. But we have terminology ready to hand that allows us to distinguish two different ways in which a sentence may have its truth value determined by a possible world. A sentence's H-intension tells us about the truth or falsity of the sentence *with respect to worlds* (with no variation in which world plays the role of the actual world). Truth with respect to possible worlds is relevant to the evaluation of '$\Box$'-modalizations. In contrast, a sentence's D-intension tells us about the truth or falsity of the sentence *at worlds considered as actual*. Truth at possible worlds considered as actual is relevant to the evaluation of '$\mathcal{F}\mathsf{A}$'-modalizations.

1.4 The simple modal conception of the two-dimensional framework

It will be clear from this brief review that Humberstone and I employed a *simple modal conception* of the two-dimensional framework. Along the second (horizontal) dimension are ranged possible worlds with respect to which sentences are evaluated in the way familiar from the semantics for standard modal languages with '$\Box$' and '$\Diamond$'. Along the first (vertical) dimension are ranged the very same possible worlds, but now playing the role of the actual world—the world with respect to which a sentence is evaluated if it occurs within the scope of the 'Actually'-operator, 'A'. Along the diagonal, the same possible worlds play both roles simultaneously.

Thus, the three one-dimensional intensions have the very same set of worlds as their domain. The domain of the D-intension might be described as 'possible worlds considered as actual', but this should not be taken to indicate a new category of world-like items. Rather, the description should be taken in an utterly flat-footed way. The truth value assigned to a sentence s for world w as argument is the truth value with respect to w (as for the H-intension) but with the same world w also playing the role of the actual world.

We can briefly contrast this simple modal conception of the framework with three others that are discussed by David Chalmers and by Robert Stalnaker: the contextual, epistemic, and metasemantic conceptions. Our conception of the first (vertical) dimension is not the *contextual* conception that is discussed and rejected by Chalmers (this volume, Chapter 4). Although there are formal similarities between modal logic and tense logic, we follow Evans in not regarding the actual world as a contextual parameter.[6] Nor is our conception the *epistemic* one favoured by Chalmers (this volume, Chapter 4). We do not build anything epistemic into the framework. Thus, Thomas Baldwin says (2001: 161): '[T]here is, in the face of it, nothing epistemological about the role of either dimension [of two-dimensional

[6] See Evans (1985). (In Davies and Humberstone (1980), this paper by Evans was incorrectly referred to as forthcoming in a Festschrift for Donald Davidson.) See also, Forbes (1983); Davies (1983). Davies and Humberstone (1980) make no attempt to extend their use of the two-dimensional framework to encompass context-dependence. In this respect, their approach is different from the approach of those whose route into the framework goes via David Kaplan's (1989) work on character and content in the semantics of demonstratives.

possible-worlds semantics].' However, although nothing epistemic is built into the framework itself, the notion of actuality does give rise to some important *a priori* truths (see below, Section 4.1).

Finally, our conception of the first dimension is not the *metasemantic* one that Stalnaker endorses (this volume, Chapter 13; see also Stalnaker 2001, 2003). The sentences that are considered in the two-dimensional framework are taken as being understood with their standard meanings.

2. Evans's Objection to the Introduction of '$\mathcal{F}$'

In his 'Comments on "Two notions of necessity" ' (this volume, Chapter 6), Evans raises a worry about Davies and Humberstone's introduction of the 'Fixedly'-operator, '$\mathcal{F}$', into a modal language. He argues that the new operator involves a quite new way of embedding sentences and that this is liable to give rise to problems.[7]

2.1 Context-shifting operators: 'A hitherto unknown form of embedding'

The analogy that Evans (1985: 357–8) draws is with a hypothetical language in which:[8]

> A sentence like 'To the left (I am hot)' as uttered by a speaker *x* at *t* is true iff there is at *t* on *x*'s left someone moderately near who is hot.

The reason why we have to recognize 'a hitherto unknown form of embedding' here is that 'the semantic value which the sentence "P(X)" ["To the left (I am hot)"] has in a context is a function of the semantic value which X ["I am hot"] would have in *another* context' (1985: 357). For consider what the semantic rule for 'To the left' must be (1985: 358; emphasis added):

> If, but apparently only if, we suppose that these operators are governed by the rule that a sentence of the form 'To the left'^(*S*) is true, *as uttered by x* at *t* iff there is someone moderately near to the left of *x* such that, *if he were to utter the sentence S* at *t*, what he would thereby say is true, we can generate the postulated truth conditions, while continuing to suppose that the only role of the first person pronoun is that of denoting the speaker.

In this case, 'To the left' is functioning as a *context-shifting operator*, just as 'As for Lloyd' would be if the sentence 'As for Lloyd (I am hot)', as uttered by Martin at *t*, were to be true just in case Lloyd is hot at *t*.

[7] Evans begins (this volume, 176): 'I confess to being a bit suspicious of the way you introduce your operator "$\mathcal{F}$", though I am quite unable to express my doubts in a compelling way.' See also Humberstone (2004), p. 29, who points out that presenting the model theory for '$\mathcal{F}$' by invoking a relation of variance between models serves to highlight the fact that, with the introduction of this operator, 'something rather new is happening'.

[8] Cf. Lewis (1980), pp. 27–8: 'To be sure, we could speak a language in which "As for you, I am hungry." is true iff "I am hungry." is true when the role of speaker is shifted from me to you—in other words, iff you are hungry. We could—but we don't.'

There is certainly a respect in which 'actually' is at least analogous to context-dependent expressions like 'I', 'here', and 'now'. For, as we ordinarily use expressions like 'actually', 'as things actually are', or 'in the actual world', these expressions take us back to how things really, *actually* actually are, even when they are embedded inside other operators. Evans stresses this point when he says (this volume, 179–80):

> You write ' "$\mathcal{F}$Aα" says: whichever world had been actual, α would have been true in the actual world.' But precisely because of the 'rigidity' of 'actual' I hear this wrong; [I] suggest you alter it to '. . . α would have been the case in that world'.[9]

So it must be acknowledged that the way in which 'A' behaves within the scope of '$\mathcal{F}$' is importantly different from the way that 'actually' behaves within the scope of other operators, including modal operators, in natural language. According to Evans, this behaviour of the 'Actually'-operator can be understood only if we regard '$\mathcal{F}$' as a context-shifting operator like 'To the left'.

Davies and Humberstone made some attempt to respond to Evans's concern that '$\mathcal{F}$' is a context-shifting operator (Davies and Humberstone 1980: 12–13; Davies 1981: 201–9). This is not the place to rehearse that attempt, but one salient claim is that the actual world should not be regarded as an aspect of context (like the speaker, time, or place of an utterance). Difference in context makes for a difference in what is said. If Lloyd and Martin both say, 'I am hot', believing what they say, then what Lloyd says and believes is not what Martin says and believes. If both on Monday and on Tuesday I say, 'Today is fine', believing what I say, then what I say and believe on Monday is not what I say and believe on Tuesday. But difference in which world is actual does not make for a difference in what is said. It is not plausible that if things had been slightly different—if a different possible world had been actual—then I would have said and believed something different in uttering 'Grass is actually green'.

2.2 Utterance difficulties

However, it is not clear that we go to the heart of Evans's suspicions about '$\mathcal{F}$' by disputing whether it is literally a context-shifting operator. For, even if the actual world is not properly an aspect of context, it might still be that the introduction of '$\mathcal{F}$' is problematic. So, what might the problematic feature be?

We have noted that the way in which the 'Actually'-operator behaves within the scope of '$\mathcal{F}$' is different from the way that 'actually' behaves within the scope of operators in natural language. But it cannot be that this difference is, by itself, a reason to find the introduction of '$\mathcal{F}$' into a formal language problematic. Evans affirms, and Davies and Humberstone deny, that '$\mathcal{F}$' must be regarded as a context-shifting operator. But Evans does not say that there is anything formally or conceptually objectionable about the introduction of a context-shifting operator. Thus (this volume, 177): 'Now I did not think and do not think that this form of embedding is incoherent, but I should like its distinctness from previously recognized forms to be made explicit.'

[9] As a result of this comment by Evans, the published version of 'Two notions of necessity' has (p. 3): ' "$\mathcal{F}$Aα" says: whichever world had been actual, α would have been true at that world considered as actual.'

What Evans does suggest is that, if '$\mathcal{F}$' is a context-shifting operator, then it may be hard to avoid 'utterance difficulties' when explaining how '$\mathcal{F}$' functions. To see how this problem arises, consider first the undisputed context-shifting operator, 'To the left', with the semantic rule (1985: 358):

> 'To the left' ^(*S*) is true, as uttered by *x* at *t* iff there is someone moderately near to the left of *x* such that, if he were to utter the sentence *S* at *t*, what he would thereby say is true.

This rule has the consequence that my utterance of 'To the left (I am speaking)' comes out true even when there is a silent person to my left. So, if '$\mathcal{F}$' is a context-shifting operator, then its semantic rule may similarly have '$\mathcal{F}$(someone actually speaks)', and '$\mathcal{F}$A(someone speaks)', come out true even though there are possible worlds in which no one speaks. This would be problematic if '$\mathcal{F}$A' is supposed to express a notion of necessity. When we say that it is contingent which possible world is actual, we surely do not have to allow that, for the corresponding notion of necessity, it is necessary that someone speaks.

This problem does not, strictly speaking, rest on the claim that '$\mathcal{F}$' is a context-shifting operator. It arises provided only that Evans is right to say that understanding '$\mathcal{F}$s' '[involves] the thought of the utterance of the embedded sentence in other circumstances' (this volume, 178). But it is not really clear why we have to accept that idea. What is clear is that, when we consider a sentence embedded within the scope of '$\mathcal{F}$' or '$\mathcal{F}$A', it will not do to consider the truth of the embedded sentence *with respect to* worlds. For truth with respect to possible worlds is relevant to understanding only '$\Box$'-modalizations. But there is an alternative to considering the truth of sentences with respect to worlds. We can consider the truth of *sentences at worlds considered as actual*. If understanding '$\mathcal{F}$A'-modalizations does not require consideration of *utterances* of the embedded sentence, then it is difficult to see why '$\mathcal{F}$(someone actually speaks)' should come out true.

In response to this, a critic might concede one point but hold to another. The critic might concede that appeal to the truth of sentences at worlds considered as actual would permit the introduction of a primitive modal operator expressing truth on the diagonal. But the critic might still maintain that the introduction of '$\mathcal{F}$' does involve utterance difficulties. It is not clear what the motivation for this position would be and I shall proceed on the provisional assumption that we can introduce '$\mathcal{F}$' without running into utterance difficulties.[10] But if this imagined critic's position were shown to be correct then we could simply forgo '$\mathcal{F}$' and introduce a primitive modal operator, '$\mathcal{D}$', for truth on the diagonal.[11] Indeed, it is of some interest to note that, at the beginning of his comments, Evans suggests the introduction of a primitive operator equivalent to the combination '$\mathcal{F}\Box$'. He may well have favoured the introduction of

[10] My brief discussion here does not respond to every aspect of what Evans says about utterance difficulties. See again Evans, this volume, esp. n. 5 and the associated text.

[11] '$\mathcal{D}$s' is true with respect to w_j with w_i playing the role of the actual world just in case, for every world w, the embedded sentence s is true with respect to w, with w also playing the role of the actual world. '$\mathcal{D}$' is thus the third of the four operators listed in the second paragraph of note 4 in Davies and Humberstone (1980). The combination '$\mathcal{F}\Box$' is equivalent to the fourth of those operators.

'$\mathcal{D}$' rather than '$\mathcal{F}$' for the same reason; namely that it 'is closer to a necessity operator right from the start' (this volume, 176).

The logic of '$\mathcal{D}$' would, of course, be different from the logic of '$\mathcal{F}$'; for example, '$\mathcal{D}$A*s*' is equivalent to '$\mathcal{D}$*s*' although '$\mathcal{F}$A*s*' is not equivalent to '$\mathcal{F}$*s*'. Truth on the vertical, previously expressed by '$\mathcal{F}$', would no longer be expressible; in particular, while '$\mathcal{D}$' is definable in terms of '$\mathcal{F}$' and 'A', '$\mathcal{F}$' is not definable in terms of '$\mathcal{D}$' and 'A'. But perhaps this would be no great loss since truth on the vertical does not correspond to any intuitive notion of necessity. And the necessity previously expressed by '$\mathcal{F}\Box$' (or equivalently by '$\Box\mathcal{F}$'), truth everywhere in the two-dimensional matrix, would now be expressed by '$\mathcal{D}\Box$' (but not by '$\Box\mathcal{D}$', which is equivalent to '$\mathcal{D}$' by itself).

3. Superficial versus Deep Contingency and Necessity

Davies and Humberstone's (1980) two notions of necessity were the necessity expressed by the familiar modal operator '$\Box$' and the necessity expressed by the novel operator '$\mathcal{F}$A'. Since the first is truth on the horizontal and the second is truth on the diagonal, let us say that the first notion is *H-necessity* and the second is *D-necessity*. Davies and Humberstone suggested that H-necessity is Evans's superficial necessity while D-necessity coincides with Evans's deep necessity. But when Evans introduced his distinction between superficial and deep contingency, he certainly did not treat it as a distinction between two modal operators in two-dimensional modal logic.

Evans characterizes superficial contingency as a property of a sentence that 'depends upon how it embeds inside the scope of modal operators'—the standard modal operators, '$\Box$' and '$\Diamond$' (1979: 179). So the identification of superficial necessity with H-necessity, the necessity expressed by '$\Box$', is straightforward. But he does not characterize deep contingency in terms of modal operators at all.

3.1 Evans on deep contingency and necessity

Deep contingency is introduced thus: 'Whether a statement is deeply contingent depends upon *what makes it true*' (1979: 179; emphasis added). By way of elucidation of this characterization, Evans tells us that 'there is an ineliminable modal element in the notion of what makes a sentence true' (206). To say that a state of affairs *makes a sentence true* is to say that, had that state of affairs obtained, the sentence would have been true. But there is also an additional constraint on the notion of making true; namely, that s and 'A*s*' are made true by the same states of affairs. They are either both deeply contingent or both deeply necessary.[12]

If we think of a sentence's being *made true by* a state of affairs along the lines of the sentence's being *true with respect to* a possible world, then this additional constraint is bound to seem puzzling. In general, s and 'A*s*' are true with respect to different possible worlds. That is why it may be that '$\Box$*s*' is false even though '$\Box$(A*s*)' is true.

[12] This additional constraint follows from two claims. First, 'A*s*' and s are 'epistemically equivalent' (1979: 210). Second, if two sentences are epistemically equivalent then they are made true by the same states of affairs (p. 205).

So how could s and 'A*s*' be made true by the same states of affairs? The way out of this apparent puzzle is to observe that Evans insists that we distinguish between truth *with respect to* a world and truth *in* a world (p. 188, n. 17). Truth with respect to possible worlds is relevant to the evaluation of '□'-modalizations and so it belongs with the notions of superficial contingency and necessity. But the notions of deep contingency and necessity go along with truth *in* possible worlds. A sentence is deeply necessary just in case it is true *in* every possible world.[13] Truth *in* a world w is glossed as: if w were to obtain, or were to be actual, then _____ would be true (p. 207). And it is subject to the constraint that s and 'A*s*' are true in the same worlds.

Truth with respect to possible worlds is, Evans says, a notion that is 'purely internal to the semantic theory' (p. 207); its role is just to deliver the correct truth values for modal sentences containing '□' and '◇'. Superficial contingency and necessity are a matter of the *properties* (specifically the truth values) *of modal sentences*. In contrast, deep contingency and necessity are a matter of what makes a sentence true and of truth *in* possible worlds. There is a modal element in this notion, but that does not mean that deep contingency and necessity are themselves fundamentally a matter of the properties of modal sentences. Rather, they are a matter of the *modal properties of* (non-modal) *sentences*. We can represent the clusters of notions associated with superficial contingency and necessity, on the one hand, and with deep contingency and necessity, on the other hand, in the following table.

Clusters of notions associated with superficial and deep contingency and necessity

Superficial	**Deep**
Truth with respect to worlds	Truth in worlds (being made true)
Purely internal to semantic theory	Not purely internal to semantic theory
Properties of modal sentences	Modal properties of sentences

Evans's final explanation of deep contingency is this (1979: 212):

> If a deeply contingent statement is true, there will exist some state of affairs of which we can say both that had it not existed the sentence would not have been true, and that it might not have existed. The truth of the sentence will thus depend upon some contingent feature of reality.

Correspondingly, a deeply necessary sentence is one whose truth depends on no contingent feature of reality. Whichever state of affairs were to obtain, whichever possible

[13] Evans stresses that, when he talks about truth in a world, he is not concerned with 'the truth of a sentence identified merely as a sequence of expression types', but with a sentence being true '*as a sentence of English*' (p. 207; italics in original). David Chalmers has pointed out that, if a sentence's truth in w as a sentence of English requires that the English language should exist in w, then this seems to make the existence of English itself deeply necessary. But this is not clearly an objection to Evans's account, so long as it is only the *abstract* language whose existence is deeply necessary. It would be problematic if the account had the consequence that it is deeply necessary that English should exist as a language *in use* or that English should be *spoken*. But that problem is avoided so long as truth in a world is not glossed in terms of truth if uttered as a sentence of English in that world. See below, Section 3.2.

world were to be actual, the sentence would still be true. A deeply necessary sentence is true *no matter what*.

It might seem at first that Evans's notions of deep contingency and necessity are technical and recherché by comparison with the notions of contingency and necessity associated with the familiar modal operators. But this is not so. Indeed, once an 'Actually'-operator is introduced, it is the idea of '$\Box$' as capturing an intuitive notion of necessary truth for sentences that stands in need of defence. In contrast, the idea that a necessarily true sentence is one that is true no matter what strikes us immediately as being right.

3.2 Deep necessity, absolute truth, and utterances

We need to say a little more about why the notion of truth *in* a world has a life of its own, rather than being purely internal to a semantic theory that specifies the truth conditions of modal sentences. A sentence is *true in a world* just in case, if that world were actual, the sentence would be *true*. This notion of truth *simpliciter*, or absolute truth, is the familiar and philosophically fundamental notion of truth as the normative end of assertion and judgement. So, there is a close conceptual connection between the notions of deep necessity, being made true by a state of affairs, and truth in a world, on the one hand, and the truth of assertions or utterances and the correctness of judgements or thoughts, on the other.

Given this close connection, it might seem natural to move to the idea that truth in a world, or being made true by a state of affairs, should be glossed directly in terms of the truth of utterances or the correctness of thoughts. So, can we say, for example, that a sentence is made true by a state of affairs just in case an utterance of the sentence in such a state of affairs would be a true utterance? Can we say that a sentence is true in a world just in case a thought in that world with the content that is conventionally expressed by the sentence would be a correct thought?

Consider an account of the truth of a sentence, s, *in* a world, w, along the lines of:

(U) If w were to be actual, then an utterance of s in w would be true.

For a wide range of cases, this gets the right results; and it is faithful to the requirement that s and 'As' should be true *in* the same worlds. In a world where grass is orange, an utterance of 'Grass is orange' or of 'Grass is actually orange' would be a true utterance. But any gloss of a sentence's truth in a world that proceeds directly in terms of utterances runs into trouble over sentences such as 'All is silent' or 'Someone speaks'.[14] Similarly, a gloss that proceeds directly in terms of having a thought with the content that would be conventionally expressed by s runs into trouble over 'No thought is going on' or 'Someone thinks'. We certainly do not want the consequence that the sentences 'I speak' and 'I think' are made true by every state of affairs, are true in every possible world, and so are deeply necessary.

[14] Strictly speaking, (U) will have every sentence that is not uttered in w come out vacuously true in w. However, if it is construed so that we consider, not w itself, but a world differing minimally from w so as to allow for the utterance of s, then 'All is silent' comes out false in w, while 'Someone speaks' comes out true in w.

Evidently, the putative principle (U) overplays the connection between the truth of sentences and the truth of utterances (or the correctness of thoughts). We must find a way to acknowledge the connection between truth and assertion without ending up with an explanation of deep necessity directly in terms of the truth of utterances. We can achieve this by linking the truth of utterances with the truth of sentences in a world through a principle such as:

> If u is an utterance of sentence s in world w, then u is a true utterance in w just in case s is true *in* w.

(See Davies and Humberstone (1980), pp. 15–17.) Given such a link, we can then retain Evans's account of the truth of a sentence, s, in a world, w:

> If w were to be actual, then s would be true.

Here, there is no mention of assertion or utterances.

3.3 Deep necessity and D-necessity

The proposal that deep necessity coincides with D-necessity or truth on the diagonal is, in essence, the proposal that Evans's notion of truth *in* a world coincides with Davies and Humberstone's notion of truth *at a world considered as actual*.

It is important that what is being suggested here is *not* that the fundamental explanation of truth in a world should be in terms of truth at a world considered as actual. That suggestion would fly in the face of the contrast that Evans draws between superficial and deep necessity. Evans says that superficial necessity is explained in terms of a theory-internal notion of truth while deep necessity is not. But, in two-dimensional possible-worlds semantics, $\text{truth}_{w,w}$—that is, truth at a world considered as actual—is a theory-internal notion that figures in the evaluation of '$\mathcal{F}$A'-modalizations just as, in one-dimensional possible-worlds semantics, truth_w is a theory-internal notion that figures in the evaluation of '□'-modalizations.

The suggestion is, rather, that the reason why the sentences that are deeply necessary turn out to be the sentences whose '$\mathcal{F}$A'-modalizations are true is that the model-theoretic notion of $\text{truth}_{w,w}$ corresponds to the notion of absolute truth—the truth at which assertion and judgement aim. Quite generally, we must be able to connect truth with validity.[15] So absolute truth must correspond to some model-theoretic notion and, given that s and 'A' are to be true in the same worlds, $\text{truth}_{w,w}$ is the only candidate.

Thus, Davies and Humberstone argue that, in the two-dimensional framework, it is with $\text{truth}_{w,w}$—rather than with truth_{w_i,w_j} or with $\text{truth}_{w^*,w}$—that absolute truth is most closely connected. Suppose, for example, that s means that grass is orange and

[15] In the case of one-dimensional possible-worlds semantics, Evans says that we must 'be able to regard absolute truth as a special case of [the theory-internal notion] truth_w' (1979: 203). In particular, if w* is the actual world then absolute truth—the truth at which assertion and judgement aim—must coincide with the specific theory-internal notion truth_{w*}: 'Only if there is this connection between the concepts will it follow from the fact that a sentence is (absolutely) true, that there is a world with respect to which it is true' (1979: 203).

consider a possible world, w, in which grass is indeed orange. If w were actual, if the state of affairs of grass's being orange were to obtain, then sentence s would be true in the absolute sense; so sentence s is true *in* w. Because s and 'As' are to be made true by the same states of affairs, 'As' must also be true *in* w. But 'As' comes out false with respect to w (or any other world) if the 'Actually'-operator is interpreted as taking us back to the real actual world, w*, where grass is green. Thus, truth *in* w does not coincide with $\text{truth}_{w^*,w}$, for example, but with $\text{truth}_{w,w}$. As Davies and Humberstone put it, 'the truth which matters, the truth at which sincere asserters in w aim, is $\text{truth}_{w,w}$' (1980: 16).

In this section, we have revisited Davies and Humberstone's suggestion that Evans's distinction between superficial and deep necessity can be rendered by the distinction between two operators in two-dimensional modal logic, '$\Box$' and '$\mathcal{F}\mathsf{A}$'. Earlier (Section 2.2) we argued that, despite worries that Evans raised, the notion of D-necessity expressed by '$\mathcal{F}\mathsf{A}$' is not subject to utterance difficulties. But suppose that someone remains unpersuaded by those arguments. If the worries are specific to the introduction of '$\mathcal{F}$' then we have offered a primitive modal operator, '$\mathcal{D}$', for D-necessity or truth on the diagonal. But perhaps it is thought that '$\mathcal{D}$' is, itself, beset by utterance difficulties; or perhaps there are residual concerns just because the behaviour of 'A' within the scope of '$\mathcal{D}$' is different from the behaviour of 'actually' within the scope of natural-language operators. A sceptic about both '$\mathcal{F}\mathsf{A}$' and '$\mathcal{D}$' can take a step back from two-dimensional modal logic to two-dimensional semantics and still accept the core of Davies and Humberstone's suggestion. Superficial necessity is H-necessity or truth on the horizontal; deep necessity coincides with D-necessity or truth on the diagonal. But, according to this sceptic, while superficial necessity is expressed by '$\Box$', deep necessity is (surprising as this may sound) not expressed by any modal operator at all.

4. Actuality and the *A Priori*

Although we are concerned with the puzzles of the contingent *a priori* and the necessary *a posteriori*, epistemological notions have been strikingly absent from the discussion up to this point. However, while nothing epistemic has been built into the two-dimensional framework itself, the notion of actuality does give rise to some important *a priori* truths.

4.1 The epistemic equivalence of s and 'As'

Evans says that the two sentences, s and 'As', are 'epistemically equivalent' (1979: 210), where epistemic equivalence is a tighter relationship than *a priori* equivalence and is explained as follows (1979: 200):

> [I]f two sentences have the same content, then what is believed by one who understands and accepts the one sentence as true is the same as what is believed by one who understands and accepts the other sentence as true. On this, very strict, view of sameness of content, if two sentences have the same content, and a person understands both, then he cannot believe what one sentence says and disbelieve what the other sentence says. When two sentences meet this condition, I shall say that they are epistemically equivalent.

The epistemic equivalence of 'As' and s (perhaps 'cognitive equivalence' would be a better term) has an important consequence. Someone who understands 'A' and s is in a position to know *a priori* that the sentence 'As' is true just in case the embedded sentence s is true and to know *a priori* that the sentence 'As $\leftrightarrow$ s' is true.

Transposing this idea into the material mode, we say that someone who understands the notion of actuality is thereby in a position to know *a priori* that, for example, the earth actually moves just in case the earth moves. Indeed, the thought that the earth actually moves and the thought that the earth moves are epistemically and cognitively equivalent. So, if it is knowable only *a posteriori* that the earth moves then equally it is knowable only *a posteriori* that the earth actually moves. And, returning to the formal mode, if it is knowable only *a posteriori* that the sentence s is true then equally it is knowable only *a posteriori* that the sentence 'As' is true.

The sentence 'As' is *a posteriori* true while 'As $\leftrightarrow$ s' is *a priori* true. Now consider the modal properties of 'As'. It is true on the horizontal and so '$\Box(\mathsf{A}$s)' is true; but it is not true on the diagonal and so '$\mathcal{F}\mathsf{A}(\mathsf{A}$s)' is false (since s is contingently true). The sentence 'As' is H-necessary and so superficially necessary; but it is D-contingent and so deeply contingent. In short, 'As' is a simple example (the simplest example) of the superficially necessary but deeply contingent *a posteriori*.

If we consider the modal properties of 'As $\leftrightarrow$ s' we find the opposite profile. It is true on the diagonal and so '$\mathcal{F}\mathsf{A}(\mathsf{A}$s $\leftrightarrow$ s)' is true; but it is not true on the horizontal and so '$\Box(\mathsf{A}$s $\leftrightarrow$ s)' is false (since s is contingently true). The sentence 'As $\leftrightarrow$ s' is D-necessary and so deeply necessary; but it is H-contingent and so superficially contingent. Thus, 'As $\leftrightarrow$ s' is a simple example (the simplest example) of the superficially contingent but deeply necessary *a priori*.

Over the very limited domain of these ur-examples, *a priority* dissociates in both directions from superficial necessity and coincides with deep necessity. But it is a further question whether there is any more general relationship between *a priority* and deep necessity. There is nothing in the two-dimensional framework itself to suggest that *a priority* should coincide with truth on the diagonal.

4.2 Is the deeply contingent *a priori* intolerable?

Evans says that 'there is nothing particularly perplexing about the existence of a statement which is both knowable *a priori* and *superficially* contingent' but that 'it would be *intolerable* for there to be a statement which is both knowable *a priori* and *deeply* contingent' (1979: 179; emphasis added). He does not provide very much in the way of argument for the claim that the combination of deep contingency with *a priority* is intolerable. But it is clear what such an argument would need to show; namely, that if the truth of an understood sentence can be known *a priori* then that truth cannot depend on any contingent feature of reality. Here we face two problems. First, we can already predict certain kinds of counterexample to the claim that what is knowable *a priori* is deeply necessary. Second, while a powerful intuition speaks in favour of some hedged version of the claim that *a priority* entails deep necessity, it is not easy to see how to provide the intuition with illuminating argumentative support, even if those predictable kinds of counterexample could be set to one side.

To see how the first problem arises, consider that we are sometimes entitled to ignore the possibility of empirical conditions that would defeat a claim to knowledge. Thus, for example, in the case of my *a posteriori* knowledge, based on perception, that I have hands, I am entitled to ignore the possibility that I am a handless brain in a vat who is the victim of a powerful but deceptive scientist (Pryor, 2000). Evidence that I am a brain in a vat would remove my epistemic warrant for believing that I have hands. But in the absence of such evidence, I can know that I have hands without taking any positive steps to rule out the brain-in-a-vat possibility.

We sometimes presume upon the non-obtaining of various empirical defeating conditions in the case of *a priori* knowledge, too. Even though a justification is empirically defeasible, it can still be an *a priori* justification provided that we are entitled simply to ignore the possibility that the empirical defeating condition obtains. For example, in following a mathematical proof, we are entitled to ignore the possibility that memory failure prevents us from keeping track of the preceding steps (Burge, 1993). In this case, (a), the proof constitutes a conclusive, and not just a *prima facie*, justification for the mathematical belief. But evidence of memory failure would threaten our justification for believing that what is before us is a proof. In other cases, (b), of *a priori* knowledge, a defeating condition would count against there being any such thing to think as the proposition whose truth we are investigating. If the defeating condition were to obtain then our putative or 'essayed' thought would not have a truth-evaluable content at all; there would be an illusion of understanding. But we are entitled to ignore the possibility that the defeating condition obtains.[16] Perhaps there are even cases, (c), of *a priori* knowledge in which we are entitled to ignore a possible defeating condition whose obtaining would be straightforwardly sufficient for the falsity of the believed proposition.[17]

In all three kinds of case of empirically defeasible *a priority*, it is utterly contingent whether the defeating condition obtains or not. But there is an important difference between cases of kind (a) and cases of the other two kinds. In cases of kind (a), provided that we do have a conclusive *a priori* justification for the mathematical belief, it is natural to maintain that the proposition believed is true as a matter of necessity. But, in cases of kinds (b) and (c), we clearly cannot move directly from *a priority* to truth no matter what. For, in those kinds of case, we presume upon contingent states of affairs (the non-obtaining of certain potential defeating conditions) that are crucial to the truth, or even to the truth-evaluability, of the proposition in

[16] Burge says (1988: 653): 'It is uncontroversial that the conditions for thinking a certain thought must be presupposed in the thinking.'

[17] Field (1996) distinguishes between *weak a priority*, which admits of empirical defeat, and *strong a priority*, which does not. He also distinguishes between primary and secondary undermining evidence, where 'secondary undermining evidence does not primarily go against the claim being undermined but against the claim that we knew it a priori' (p. 362). Field's final account of strong *a priority* is that it does not admit of primary empirical defeat. So cases of kind (a) could still be cases of strong *a priority*, while cases of kinds (b) and (c) could only be cases of weak *a priority*.

See also Peacocke (2004: 24–31) for a similar distinction (p. 30) between defeasibility of identification (cf. Field's secondary undermining evidence) and defeasibility of grounds (cf. Field's primary undermining evidence) and for the important notion of relative *a priority* (p. 26).

question. There are ways in which our thought could be false, or ways in which our putative thought might not even be truth-evaluable, that are not ruled out by our *a priori* justification. So, even given an intuition to the effect that what can be established *a priori* cannot depend on any contingent feature of reality, the most that we could reasonably conclude would be that the proposition is true in all those worlds that include the presumed-upon states of affairs.

Let us turn now to the second problem. Even if we set aside the phenomenon of empirically defeasible *a priori* justification found in cases of kinds (b) and (c), it is difficult to provide illuminating argumentative support for the claim that *a priority* entails deep necessity. Suppose for *reductio* that the truth of some understood sentence, s, can be known *a priori* although s is deeply contingent. (And suppose that this deep contingency is not just a reflection of the fact that the possibility of certain empirical defeating conditions is legitimately ignored in the course of the *a priori* justification.) The truth of s depends on the obtaining of a contingent state of affairs, S. *A priori* knowledge that s is true would provide an *a priori* guarantee that S does indeed obtain. But, even if S does obtain, still it might not have obtained. It is not guaranteed to obtain. As Evans puts it (1979: 212): 'A deeply contingent statement is one for which there is no guarantee that there exists a verifying state of affairs.' It might seem a very short step from this point to a contradiction: S is not guaranteed to obtain even though we have a guarantee that S does obtain. But this short step involves a slip between modal and epistemic notions of guarantee: S is not *modally* guaranteed to obtain even though we have an *epistemic* guarantee that S does obtain. So, instead of saying that because S is contingent it is not guaranteed to obtain, it would be better to stress that contingency is a modal notion: S modally might not have obtained. In order to complete the *reductio*, we would then need to show that if S modally might not have obtained then we cannot be *a priori* epistemically guaranteed that S does obtain. But this is uncomfortably close to what we were supposed to be showing in the first place, namely, that *a priority* entails deep necessity.[18]

What the attempted *reductio* does achieve is a shift from a claim about a sentence to a claim about a state of affairs. This serves to highlight the very close relationship between deep necessity for sentences and the necessary obtaining of states of affairs. There is a much looser relationship between superficial necessity for sentences and the necessary obtaining of states of affairs.

To see this contrast, consider again '**A**s ↔ s' as a simple example of the *superficially* contingent *a priori*. The superficial contingency of this sentence depends on the occurrence of the 'Actually'-operator. But '**A**s' and s are made true by the same states of affairs. So the superficially contingent sentence '**A**s ↔ s' is made true by the same states of affairs as the sentence 's ↔ s', which is superficially (and deeply) necessary. Similarly, the superficially necessary sentence '**A**s' is made true by the same states of affairs as the sentence s, which is superficially (and deeply) contingent. So we must not conflate superficial contingency as a property of sentences with contingency as a

[18] See Forbes (1989: 152) for a similar argument that 'the natural way of trying to show that everything contingent is *a posteriori*' breaks down because it 'assumes what it is supposed to be establishing'.

property of states of affairs. The sentence '$\mathsf{A}s \leftrightarrow s$' is superficially contingent, but it is made true by a state of affairs that obtains as a matter of necessity. The sentence '$\mathsf{A}s$' is superficially necessary, but it is made true by a state of affairs that modally might not have obtained.

Deep contingency, in contrast with superficial contingency, is defined in terms of making true and thus cannot depend on the pattern of occurrences of the 'Actually'-operator. So we can safely move between deep contingency as a property of sentences and contingency as a property of states of affairs. A deeply contingent sentence is made true by a state of affairs that might not have obtained. As we have seen, the question whether a sentence could be *a priori* true yet deeply contingent then becomes the question whether we could have an *a priori* epistemic guarantee that a state of affairs obtains even though that state of affairs modally might not have obtained. A negative answer to this question is supported by intuition, rather than by independent argument.

To the extent that the combination of *a priority* and *contingency for states of affairs* is intolerable, the combination of *a priority* and *deep contingency for sentences* is equally intolerable. But, however it may be with states of affairs, the combination of *a priority* and *superficial* contingency for sentences may be both tolerable and unperplexing, as the example of '$\mathsf{A}s \leftrightarrow s$' shows.[19]

4.3 Sense, reference, and asymmetry

The idea we have reached is that there is a very close connection between the deep modal properties of sentences and the modal properties of states of affairs. One way of developing this idea would be to follow Graeme Forbes (1989) in assigning a state of affairs (rather than, as Frege would have it, a truth value) to each sentence as its referent or *Bedeutung*.[20] A true sentence is deeply contingent if its referent is a state of affairs that might not have obtained. A sentence is deeply necessary if its referent is a state of affairs that obtains as a matter of necessity; that is, a state of affairs that would obtain no matter which world were to be actual. *Deep modal properties* can then be described as belonging fundamentally at *the level of reference* and a sentence operator expressing a deep modal property can be properly classified as *extensional*. If s_1 and s_2 have the same referent then 'It is deeply contingent that'$^\wedge s_1$ is true just in case 'It is deeply contingent that'$^\wedge s_2$ is true.

A sentence has not only a referent, but also a sense; namely, the thought—perhaps better, the thought content—that it expresses (Frege, 1892). Superficial modal properties cannot be transposed from sentences to their senses because (as in the case of

[19] In *Naming and Necessity*, before introducing his apparent examples of the contingent *a priori* and the necessary *a posteriori*, Saul Kripke makes some suggestions about why 'people have thought that these two things ["necessary" and "*a priori*"] must mean the same' (1980: 38). Concerning the move from *a priority* to necessity, he says (1980: 38): 'I guess it's thought that . . . if something is known *a priori* it must be necessary, because it was known without looking at the world. If it depended on some contingent feature of the actual world, how could you know it without looking? Maybe the actual world is one of the possible worlds in which it would have been false.'

[20] Forbes (1989), ch. 5. For this purpose, states of affairs are abstract state types that might or might not obtain. Cf. Barwise and Perry (1983); Taylor (1976, 1985).

'As' and s) a sentence that is superficially necessary and a sentence that is superficially contingent may express the same thought content. But deep modal properties can be transposed from sentences to their senses. Evans says that epistemically equivalent sentences are made true by the same states of affairs (1979: 205). So sentences that express the same thought content never differ in their deep modal properties. In Forbes's framework, the point follows from an instance of the Fregean doctrine that sense determines reference. If two sentences have the same sense—that is, express the same thought content—then they are assigned the same state of affairs as their referent, and so they have the same deep modal properties.

While a sentence operator expressing a deep modal property is extensional, propositional attitude operators are classified as *intensional*. For it is not, in general, correct that if s_1 and s_2 have the same referent then 'Ralph believes that'^s_1 is true just in case 'Ralph believes that'^s_2 is true (Forbes, 1989, 121). Thought contents are discriminated more finely than states of affairs and being believed by Ralph is fundamentally a property of thought contents.

Epistemic notions such as *a priority* are like propositional attitude notions, and unlike deep modal notions, in belonging fundamentally at *the level of sense*. Thus, when we say that an understood sentence is *a priori* true we mean something along the following lines. Just in virtue of grasping the thought content that the sentence expresses (where this includes grasping the concepts that are constituents of that content), a subject is in a position to know that the thought content is correct. This *a priori* knowledge that the thought content is correct furnishes an *a priori* epistemic guarantee that the state of affairs that is the referent of the understood sentence obtains.

As we have seen (Section 4.2), it may be that this epistemic guarantee is furnished only against a background of presumed-upon conditions. So the intuition that *a priority* entails necessity for states of affairs—powerful as it may be—must be hedged. If an understood sentence is *a priori* true then the state of affairs that is the referent of the sentence obtains in all those possible worlds in which the presumed-upon conditions also obtain. Taking into account both the hedge and the lack of independent argumentative support, we said that, to the extent that the combination of *a priority* and contingency for states of affairs is intolerable, the combination of *a priority* and deep contingency for sentences is equally intolerable. We have now set the close connection between the modal properties of states of affairs and the deep modal properties of sentences in Forbes's Fregean framework of sentence, sense, and reference. The point of doing this is not to provide any new argument in support of the claim that *a priority* entails deep necessity. The point is, rather, to shed some light on the question whether intuitive support for the claim that *a priority* entails deep necessity carries over to the converse claim that deep necessity entails *a priority*.[21]

[21] Concerning the move from necessity to *a priority*, Kripke (1980: 38) credits people with the following thought: '[I]f something not only happens to be true in the actual world but is also true in all possible worlds, then, of course, just by running through all the possible worlds in our heads, we ought to be able with enough effort to see, if a statement is necessary, that it is necessary, and thus know it *a priori*.' But he immediately continues that 'really this is not so obviously feasible at all'.

With the issues set in a Fregean framework, we see that the inference from *a priority* to deep necessity involves a shift from a notion that belongs at the level of sense to a notion that belongs at the level of reference. In general, sense and reference are asymmetrically related. Sense determines reference, but there is no route back from reference to sense. So it is not unreasonable to suppose that *a priority* and deep necessity are also asymmetrically related.

It may be, for example, that many different thought contents are modes of presentation of a single state of affairs, S. Suppose that, for one of these thought contents, M, just grasping M is sufficient to put a subject into a position to know that the thought content is correct. From this, we infer that S obtains as a matter of necessity—or, at least, that S obtains in all those worlds in which certain presumed-upon conditions obtain. This is the plausible move from sense to reference. But, even if S obtains in all possible worlds, it would surely be hasty to move back from reference to sense. So we should not infer that, for every other thought content, M′, that is also a mode of presentation of S, just grasping M′ puts a subject into a position to know that the thought content is correct.

Summary: Over a domain of ur-examples, 'As $\leftrightarrow$ s' and 'As', *a priority* coincides with deep necessity and so with D-necessity or truth on the diagonal. But there is nothing in the simple modal conception of the two-dimensional framework to suggest that *a priority* should always coincide with truth on the diagonal. There is a powerful intuition that seems to support some version of the claim that *a priority* entails deep necessity. But, first, the claim must be hedged and, second, it is difficult to provide the intuition with illuminating argumentative support. More importantly, the relationship between *a priority* and deep necessity appears to be asymmetric. The inference from *a priority* to deep necessity involves a shift from sense to reference. So, for general Fregean reasons, we should not expect that intuitive support for that inference would carry over to the converse inference from deep necessity to *a priority*, since this involves a shift from reference back to sense.

5. *A Priority*, Deep Necessity, and Descriptive Names

We have seen that the relationship between *a priority* and deep necessity is complicated by the phenomenon of empirically defeasible *a priori* justification. But if we set this complication aside then there is a powerful intuition that *a priority* entails deep necessity and there is, in the sentence 'As $\leftrightarrow$ s', an ur-example of the superficially contingent but deeply necessary *a priori*. So an obvious strategy for understanding an apparent example of a sentence that is contingent and *a priori* is to show that the sentence plays some more or less complex variation on the theme of 'As $\leftrightarrow$ s'.

5.1 The contingent *a priori* and descriptive names

The sentence:

(1) If anyone uniquely invented the zip then the actual inventor of the zip invented the zip.

with the formulation:

(2) $(\exists x)(x$ uniquely invented the zip$) \rightarrow$
[The x: $\mathsf{A}(x$ invented the zip$)]$ (x invented the zip).[22]

is an apparent example of the contingent *a priori*. Sentence (2) is *a priori* true because it is epistemically equivalent to the obviously *a priori*:

(3) $(\exists x)(x$ uniquely invented the zip$) \rightarrow$
[The x : x invented the zip] (x invented the zip).

Sentence (2) is superficially contingent since its '$\Box$'-modalization is false and the sentence:

$\Diamond$ $((\exists x)(x$ uniquely invented the zip) &
$\sim$[The x: $\mathsf{A}(x$ invented the zip$)]$ (x invented the zip))

is true. It is surely possible that someone other than Whitcomb L. Judson, the person who actually invented the zip, should have invented the zip. Presumably, it is possible that Tiny Tim should have uniquely invented the zip.

But sentence (2) is deeply necessary. Because (2) is epistemically equivalent to (3) it is made true by the same states of affairs as (3)—a sentence whose 2D-intension is everywhere true. Another way to see that (2) is deeply necessary is to observe that (3) differs from (2) only by the removal of the single occurrence of the 'Actually'-operator, so that (2) and (3) have the same D-intension.[23] Sentence (3) is certainly D-necessary; so (2) is also D-necessary, and thus deeply necessary.

This combination of properties—*a priori*, superficially contingent, deeply necessary—is just that Evans (1979: 193) claims for:

(4) If anyone uniquely invented the zip then Julius invented the zip.

Here 'Julius' is a *descriptive name* whose reference is fixed by the description 'the inventor of the zip'. Evans supposes that 'Julius' is introduced into the language by the stipulation (1979: 181):

(D) Let us use 'Julius' to refer to whoever invented the zip.

And he restricts attention to the initial period of the name's use, when it is 'unquestionably a ''one-criterion'' name' (1979: 181). This restriction is crucial to Evans's account of descriptive names. A name that is originally introduced by way of a reference-fixing description may evolve into an ordinary proper name and the conditions for understanding an ordinary proper name of the inventor of the zip are quite different from the conditions for understanding the descriptive name 'Julius'.[24]

[22] The consequent of this conditional regiments the definite description by using the notation of restricted quantification. An alternative Russellian version of the consequent would be:

$(\exists x)(\mathsf{A}(x$ invented the zip$)$ & $(\forall y)(\mathsf{A}(y$ invented the zip$) \rightarrow (y = x$ & y invented the zip$)))$.

[23] For any sentence, α, that is free of '$\Box$' and '$\mathcal{F}$', if α' results from α by removal of all occurrences of 'A', then α and α' have the same D-intension.

[24] See Evans (1979), 180–2; Davies and Humberstone (1980), 18; Baldwin (2001), 166.

Now consider the three properties that Evans claims for sentence (4). First, (4) is *a priori* because 'someone can know that the sentence [4] is true, simply in virtue of knowledge he has as a speaker of the language' (1979: 192–3.). This is not just a matter of knowing *a priori* that (4) expresses some truth or other but not knowing what truth it expresses. Rather (p. 182):

> It is sufficient to understand 'Julius' that one know that it refers to whoever invented the zip. This knowledge can certainly be possessed whether or not there is such a person, and possessing it, one is in a position to know exactly what conditions have to be satisfied for sentences containing the name to be true, and hence to understand them.

Second, sentence (4) is superficially contingent because 'a world in which someone who did not actually invent the zip invents the zip is a world *with respect to which* the antecedent of the conditional [4] is true, but the consequent, and thus the whole conditional, is false' (p. 193; emphasis added). So the '□'-modalization of (4) is false.

But third, sentence (4) is not deeply contingent because 'there is no contingent feature on which its truth depends': it 'demands nothing of the actual world' (p. 212). Whichever world were to be actual, sentence (4) would still be true; that is, true as a sentence of English governed by the stipulation (D). Suppose, for example, that Tiny Tim had invented the zip. Then the (non-modal) sentences 'Tiny Tim invented the zip', 'Tiny Tim is Julius' and 'Julius invented the zip' would all have been true;[25] and sentence (4) would also have been true. If no one had uniquely invented the zip then (4) would still have been true. Sentence (4) is deeply necessary; it is true no matter what.[26]

In short, we can show that (4) is *a priori* true and superficially contingent, but deeply necessary, by pointing to modal and epistemic similarities between the descriptive name 'Julius' and the definite description 'the actual inventor of the zip'. Sentence (4) is thus revealed as playing a variation—much the same variation as sentence (1)—on the theme of '$\mathsf{A}s \leftrightarrow s$'.

5.2 The necessary *a posteriori* and descriptive names

Just as there is an obvious strategy for understanding an apparent example of a sentence that is contingent and *a priori*, so also there is an obvious strategy for understanding an apparent example of a sentence that is necessary and *a posteriori*. We show that the sentence plays some variation on the theme of '$\mathsf{A}s$'.

Whitcomb L. Judson invented the zip fastener. So the following sentence is *a posteriori* true and contingent:

(5) The inventor of the zip = WLJ.

[25] This is *not* to say that the *modal* sentence, 'If Tiny Tim had invented the zip then Tiny Tim would have been Julius', is true. See Evans (1979), 192.

[26] The deep necessity of sentence (4) is not a surprising result, given other aspects of Evans's account. As Evans conceives descriptive names, the belief that Julius is F (the belief expressed by 'Julius is F') is the very same belief as the belief that the inventor of the zip is F (Evans 1979: 202): 'We do not get ourselves into new belief states by "the stroke of a pen" (in Grice's phrase)—simply by introducing a name into the language'.

Consequently, the result of prefixing (5) with the 'Actually'-operator:

(6) Actually (The inventor of the zip = WLJ)

is *a posteriori* and superficially necessary, but deeply contingent. So too (provided that we ignore complications about contingent existence) is:

(7) The actual inventor of the zip = WLJ.

Sentence (7) is superficially necessary because its '□'-necessitation is true (again ignoring complications about contingent existence). It is deeply contingent because, if a world in which Tiny Tim invented the zip had been actual, then (7) would have been false.

Now consider:

(8) Julius = WLJ.

As a true identity statement involving proper names, this is an apparent example of the necessary *a posteriori*. But, the descriptive name 'Julius' is modally and epistemically similar to the 'actually'-embedding description 'The actual inventor of the zip' ('The *x* such that *x* actually invented the zip'). So sentence (8) is epistemically and modally like (7). Thus, sentence (8) plays a variation on the theme of 'As' and is an example of the superficially necessary *a posteriori*, but not of the deeply necessary *a posteriori*.

5.3 Ordinary proper names

Over the domain that includes these examples involving descriptive names, (4) and (8), in addition to the ur-examples ('$\mathsf{A}s \leftrightarrow s$') and ('$\mathsf{A}$s'), *a priority* coincides with deep necessity and so with truth on the diagonal. But let us now consider examples that involve only ordinary proper names.

Suppose first that, in our example of the superficially contingent but deeply necessary *a priori*, sentence (4), we eliminate the descriptive name 'Julius' in favour of an ordinary proper name, 'WLJ'. The result:

(9) If anyone uniquely invented the zip then WLJ invented the zip.

does not even appear to be an example of the contingent *a priori*. Like sentence (5), it is (both superficially and deeply) contingent but only *a posteriori* true. So sentence (9) does not present any threat to the coincidence of *a priority* with deep necessity.

However, suppose second that, instead of our example of the superficially necessary but deeply contingent *a posteriori*, sentence (8), we consider a true identity statement involving only ordinary proper names, such as 'Cicero = Tully' or:

(10) Slim Dusty = David Gordon Kirkpatrick.

This does still appear to be an example of the necessary *a posteriori*.

Here, I assume that the semantic contribution of an ordinary proper name is to be stated in an object-dependent way. There is a semantic connection between the name and its bearer and not just, as in the case of a descriptive name, between the

name and a descriptive condition. An ordinary proper name cannot refer to an object other than its (actual) bearer without a change in meaning. So long as its meaning is maintained, it refers to the same object both *with respect to* every possible world and *in* every possible world (again, we ignore complications about contingent existence). We might say that an ordinary proper name is both a superficially rigid, and a deeply rigid, designator.

Given this assumption, a true identity statement involving only ordinary proper names is both H-necessary and D-necessary. Indeed, its 2D-intension is everywhere true. Thus, a sentence like (10) does present a challenge to the coincidence of *a priority* with deep necessity because it threatens the claim that deep necessity entails *a priority*.

Summary: The overall situation suggested by the examples in this section (where we have set aside the complications of empirically defeasible *a priority*) is this. First, apparent examples of the contingent *a priori* and the necessary *a posteriori* that involve descriptive names present no challenge to the coincidence of *a priority* with deep necessity or truth on the diagonal. Apparent examples of the contingent *a priori* are consistent with the claim that *a priority* entails deep necessity; and apparent examples of the necessary *a posteriori* are consistent with the claim that *a posteriority* entails deep contingency; that is, that deep necessity entails *a priority*.

Second, when we replace descriptive names with ordinary proper names we do not produce even apparent examples of the contingent *a priori*. So there is still no threat to the claim that *a priority* entails deep necessity.

But, third, with ordinary proper names we produce apparent examples of the necessary *a posteriori* that are both superficially and deeply necessary. So these examples threaten the claim that deep necessity entails *a priority*. This overall situation is, of course, entirely consistent with the idea, defended in Section 4.3, that the relationship between *a priority* and deep necessity may be asymmetric. Intuitive support for the inference from *a priority* to deep necessity does not carry over to the converse inference from deep necessity to *a priority*.

6. The Descriptive Names Strategy

There is, clearly enough, a general strategy for bringing the epistemic distinction between *a priority* and *a posteriority* more fully into alignment with the modal distinction between deep necessity and deep contingency, so that *a priority* will more nearly coincide—will perhaps coincide perfectly—with truth on the diagonal. The strategy is to treat all referring expressions as being, or as being relevantly similar to, descriptive names. This is the strategy adopted, for example, by Frank Jackson (1998a, 1998b, 2004) both for natural kind terms like 'water' and for ordinary proper names of planets, places, and people.

In the case of natural kind terms, I think that the descriptive names strategy is quite plausible. In the end, I am somewhat inclined against it, but there is important work that still needs to be done on developing an alternative. In the case of ordinary proper names, however, I am more firmly inclined to reject the descriptive names strategy,

and to accept that there will be residual examples of the deeply necessary *a posteriori*. In this section, I shall briefly indicate why.[27]

6.1 Description-theoretic accounts of reference

I begin with some very familiar background. In *Naming and Necessity*, Kripke offers three kinds of argument against description-theoretic (descriptivist) accounts of the reference of ordinary proper names: semantic arguments, epistemic arguments, and modal arguments.[28]

Suppose that 'N' is a name in the language or idiolect of U. Then, according to a descriptivist account of proper names, there is a description, 'the H', such that the semantic condition for an object *x* to be the referent of 'N' (in the language or idiolect of U) is simply that *x* should be the unique H. Suppose further that a semantic theory for a language states what a speaker knows just in virtue of knowing or understanding the language: a theory of meaning is a theory of understanding. Then U understands the name 'N' (or knows its meaning) by knowing that 'N' refers to whichever object (if any) is uniquely H. This is the semantic aspect of a descriptivist account of proper names. A *semantic argument* against descriptivist accounts challenges these claims about meaning, reference and understanding.

A descriptivist account also says that a sentence containing 'N', say, 'N is F', is epistemically and cognitively equivalent (for U) to the sentence, 'The H is F'. To think that N is F is to think that the H is F. To know or discover that N is F is to know or discover that the H is F. This is the epistemic aspect of a descriptivist account of proper names, and an *epistemic argument* against descriptivist accounts challenges these claims about thought and knowledge.

A descriptivist account of proper names says one more thing. It says that a sentence containing 'N' is modally equivalent to the sentence that results by replacing 'N' with its reference-determining description, 'the H'. The modal equivalence of 'N is F' and 'The H is F' involves at least the requirement that, if the two sentences are embedded in the same modal context, then the resulting modal sentences should have the same truth value. Thus, for example, 'Necessarily, if something is uniquely H, then N is H' and 'Necessarily, if something is uniquely H, then the H is H' should have the same truth value. As a result, this third aspect of the descriptivist account is initially extremely implausible. For example, it might be proposed that the reference-determining description for the name 'Aristotle' is 'the teacher of Alexander'. But the sentence 'Necessarily, if someone uniquely taught Alexander, then Aristotle taught Alexander' is false, whereas 'Necessarily, if someone uniquely taught Alexander, then the teacher of Alexander taught Alexander' is true, or at least has a true reading.

Descriptivists usually respond to this problem by choosing 'actually'-embedding reference-determining descriptions. Certainly, the description 'the *actual* teacher of Alexander' comes closer to matching the behaviour of 'Aristotle' in modal sentences than the description 'the teacher of Aristotle' does. But, in the light of Evans's

[27] Clearly, the relationship between *a priority* and truth on the diagonal requires more extended consideration than it can receive here. See my '*A priority* and truth on the diagonal' (forthcoming).

[28] Kripke (1980). In the next few paragraphs, I closely follow Soames (2002), ch. 2.

distinction between superficial and deep modal properties, we should insist that modal equivalence is not just a matter of pairs of modal sentences having the same truth values. It is also a matter of pairs of *non-modal sentences*, 'N is F' and 'The H is F', having the same *modal properties*. A *modal argument* against descriptivist accounts challenges these claims about the truth values of modal sentences and about the modal properties of non-modal sentences.

Descriptive names have semantic, epistemic, and modal properties corresponding to the three aspects of a descriptivist account of proper names (Section 5.1). First, if 'M' is a descriptive name then its reference-fixing description, 'the G' (or the 'actually'-embedding description, 'the actual G'), plays exactly the reference-determining role for 'M' that is specified by the semantic aspect of a descriptivist account of proper names. Second, the sentence 'M is F' is epistemically and cognitively equivalent to 'The G is F' (or 'The actual G is F'). And third, 'M is F' and 'The actual G is F' are modally equivalent; they have the same modal profile. They are true *with respect to* the same possible worlds; so substitution of one for the other within the scope of the modal operators '$\Box$' and '$\Diamond$' makes no difference to truth value. And they are true *in* the same possible worlds: whichever state of affairs were to obtain, whichever world were to be actual, the sentences 'M is F' and 'The actual G is F' would be true together or false together. Because of these two aspects of modal equivalence, 'M is F' and 'The actual G is F' agree in their superficial, and in their deep, modal properties.

Clearly, then, the descriptive names strategy can be assessed in the light of the three kinds of argument that Kripke advanced.

6.2 Three arguments against the descriptive names strategy

Suppose that an advocate of the descriptive names strategy proposes that an ordinary proper name, 'N', in the language or idiolect of U, is or is relevantly similar to a descriptive name. The reference-fixing description, 'the G', must meet the condition that an object x is the referent of 'N' just in case x is uniquely G. So it is likely that the advocate of the descriptive names strategy will offer a reference-fixing description that incorporates the kinds of conditions that would be mentioned in a good theory of reference.[29]

This choice of reference-fixing description protects the descriptive names strategy from objections along the lines that the description 'the G' is liable to pick out an object that is not the referent of 'N'. But a *semantic argument* against the strategy can press on the requirement that the speaker, U, should know what the descriptive conditions on the reference of 'N' are. After all, U is supposed to understand 'N' by *knowing* that it refers to whichever object (if any) is uniquely G.

A defender of the strategy can respond to this kind of argument by appealing to the notion of *implicit* knowledge or grasp of the reference-determining condition (Jackson 1998b, 210–12 and 2004, 272–3). For the purposes of the present brief

[29] See, e.g., Jackson (1998b), esp. pp. 208–12 and (1998a), p. 40, n. 16; see also Kroon (1987) and (2004).

discussion, I shall allow that the semantic argument against the descriptive names strategy can be met in this way. Whether or not it is ultimately correct to make this concession, considerations that are problematic for the strategy emerge when we turn to the epistemic and modal arguments in the light of this response to the semantic argument.

Consider *epistemic arguments*. An advocate of the descriptive names strategy says that 'N is F' is epistemically and cognitively equivalent to 'The (actual) G is F'. Thus, to think, know, or discover that N is F is to think, know, or discover that the G is F. But this is not a compelling claim about thought contents. The description 'the (actual) G' incorporates the kinds of conditions that would be mentioned in a good theory of reference for proper names. So someone who thinks that the G is F thereby deploys concepts that figure in theories of reference. But it is not very plausible that, when an ordinary speaker, U, thinks that N is F, he or she deploys those reference-theoretic concepts. Intuitively, it seems that U might not even possess those concepts.

Jackson describes as 'a blind alley' the suggestion that a description theory of reference is to be resisted on the grounds that we are able to think about, and to use language to convey information about, *objects*. His reason for rejecting the suggestion is that 'you cannot give information about objects without giving information about their properties . . . we access objects via their properties' (1998b: 216). Now, it is surely correct that there is a sense in which my ability to think about an object depends on the properties of that object. Thus, suppose that I am able to think of a friend, Z, in virtue of having a capacity to recognize him. This recognitional capacity will be underpinned, we may assume, by a piece of information-processing machinery that is sensitive to various properties of visually presented people. This device will fire in the presence of my friend (provided that he is not in disguise). It would also fire in the presence of any person who was like my friend in respect of the properties to which the device is sensitive. So there is a description 'the K' that a person must satisfy if the device is to fire.

But none of this establishes a thesis about thought contents to the effect that, when I think *that Z is F*, I am really thinking *that the K is F*. The properties that are mentioned in the descriptive condition are implicated in the subpersonal-level whirrings and grindings of the device that underpins my recognitional capacity. They are the properties to which the device is sensitive. But no concepts of those properties need figure in my thinking. Similarly, we can accept that 'N' refers to whoever satisfies a description, 'the G', and even allow an implicit grasp of the reference-determining condition, without agreeing that, when I think that N is F, I am really thinking that the G is F.

Finally, consider *modal arguments* against the descriptive names strategy. According to the strategy, 'N is F' is supposed to be modally equivalent to 'The actual G is F'. The problem here does not flow from the first of the two requirements for modal equivalence, having to do with embedding in modal contexts. Substitution of 'The actual G is F' for 'N is F' within the scope of the modal operators '$\Box$' and '$\Diamond$' will, indeed, make no difference to truth value. But the second requirement for modal equivalence, having to do with modal properties—including deep modal properties—of non-modal sentences, is more problematic. For it is not obvious that whichever

state of affairs were to obtain, whichever world were to be actual, the non-modal sentences 'N is F' and 'The actual G is F' would be true together or false together.

To see how the problem arises, consider the name 'DBM' of David Braddon-Mitchell. Suppose that the reference-determining description for 'DBM' is something along the lines of 'the person whose properties cause so-and-so device to fire etc.'; and imagine that we do not press any epistemic argument against this proposal. Now consider a possible state of affairs in which David has a beard, but it is a beardless man, Nigel, whose properties cause so-and-so device to fire. If this state of affairs were to obtain, if such a possible world were to be actual, then the sentence:

(11) DBM is bearded.

would be true (without any change in its meaning). But the sentence:

(12) The person whose properties actually cause so-and-so device to fire etc. is bearded.

would be false. For this sentence is *made true* by the same states of affairs as 'The person whose properties cause so-and-so device to fire etc. is bearded'.

Similarly, the sentence:

(13) The properties of DBM cause so-and-so device to fire.

would be false, while:

(14) The properties of the person whose properties actually cause so-and-so device to fire etc. cause so-and-so device to fire.

would be true. Thus, 'DBM' is not a descriptive name with its reference fixed by the description 'the person whose properties cause so-and-so device to fire etc.'

This section is very far from providing a full cost-benefit analysis of the descriptive names strategy for bringing *a priority* into alignment with deep necessity or truth on the diagonal. But, provisionally, it seems to me that epistemic and modal arguments cast some doubt on the prospects for the descriptive names strategy, at least in its application to ordinary proper names. Examples of the deeply necessary *a posteriori*, such as true identity statements involving ordinary proper names, will remain.

7. Evans's Account of Descriptive Names as Referring Expressions

At many points in the last two sections, we have relied on modal and epistemic similarities between descriptive names and 'actually'-embedding definite descriptions—between 'Julius' and 'the actual inventor of the zip', for example. Because the properties that Evans claimed for the sentence:

(4) If anyone uniquely invented the zip then Julius invented the zip.

are just those of:

(1) If anyone uniquely invented the zip then the actual inventor of the zip invented the zip.

Davies and Humberstone suggested that 'a descriptive name with its reference fixed by "the G" is nothing other than a conventional abbreviation of (or at least, an expression whose sense is that of) "the actual G" ' (1980: 11). This suggestion seems to be accepted by some as an account of Evans's own views.[30] But, in his 'Comments on "Two notions of necessity" ', Evans explicitly rejected the suggestion that descriptive names are abbreviations of 'actually'-embedding descriptions (this volume, 179): 'So you would expect me to dissent from your suggestion that a descriptive name is a conventional abbreviation for a definite description embedding "actually".' In this final section of the paper, I shall address the question why Evans was so firmly against the idea that descriptive names belong semantically with definite descriptions.

7.1 Descriptive names, definite descriptions, and the reference relation

According to Evans, descriptive names have two crucial features (1979: 180):

> First, a descriptive name is a referring expression; it belongs to that category of expressions whose contribution to the truth conditions of sentences containing them is stated by means of the relation of reference. Second, there is a semantical connection between the name and a description; the sense of the name is such that an object is determined to be the referent of the name if and only if it satisfies a certain description.

This is likely to strike us, at least initially, as a surprising combination of features. For we are familiar, from Evans's work on reference, with a contrast between a genuine or 'Russellian' singular term, 'whose significance depends upon its having a referent' (1982, p. 12 and *passim*), and a definite description, whose significance can be grasped independently of whether it has a denotation. Understanding a Russellian singular term involves knowing *of* a particular object that the term refers to it; it involves having an object-dependent thought. For such an expression, merely knowing that it refers to whichever object satisfies a particular descriptive condition (if any object does) cannot suffice for understanding.

Against the background of these ideas about Russellian singular terms as examples of referring expressions, the two features that Evans associates with descriptive names may seem to be in tension. It may be tempting to think that, if 'Julius' is a referring expression, then someone who knows only that there is a semantic connection between 'Julius' and the description 'the inventor of the zip' does not understand 'Julius'. According to this tempting thought, a person who knows the stipulation:

(D) Let us use 'Julius' to refer to whoever invented the zip.

by which 'Julius' was introduced can know that 'Julius' refers to whoever invented the zip (assuming that it refers at all). But this does not amount to understanding 'Julius' because someone who knows *only* the stipulation (D) does not know *of* any individual, and in particular does not know *of* Whitcomb L. Judson, that he invented the zip and so is the referent of the singular term 'Julius'.

[30] See, e.g., Baldwin (2001), 166: 'On a semantic account of the matter, Evans ... simply introduced the term "Julius" as an abbreviation of the description "the actual inventor of the zip". This seems indeed to be the way in which Evans himself thought of the matter.'

In line with this tempting thought, it might be proposed that someone could, just in virtue of knowing the stipulation (D), know that sentence (4) expresses a truth, but would not thereby know what truth it is that the sentence expresses (see Donnellan 1977: 18). This is certainly *not* Evans's position. But Evans needs to explain why, given his account of what is involved in understanding sentence (4), he maintains that 'Julius' is a referring expression.

The key to this explanation lies in a distinctive view about reference coupled with the conception of a referring expression as any expression 'whose contribution to the truth conditions of sentences containing [it] is stated by means of the relation of reference' (Evans 1979: 184). First, it is agreed on all sides that reference is a relation. But Evans's distinctive view is that reference is just 'whatever relation it is between expressions and objects which makes the following principle true' (1979: 184):

(P) If $R(t_1 \ldots t_n)$ is atomic, and $t_1 \ldots t_n$ are referring expressions, then $R(t_1 \ldots t_n)$ is true iff $<$the referent of $t_1 \ldots$ the referent of $t_n>$ satisfies R.

No requirement of a causal relation between expression and object, for example, is built into the notion of reference.

Second, although reference is a relation, the semantic contribution of a referring expression need not be stated by simply asserting that the relation of reference obtains between the expression and some particular object. Nor must understanding a referring expression always involve an object-dependent thought. In the familiar case of a Russellian singular term, such as an ordinary proper name, the semantic contribution will be stated in an object-dependent way, along the lines of:

(15) The referent of 'John' = John.

But it is equally the relation of reference that is at work in the clause:

(16) $(\forall x)$ (Refers to ('Julius', x) $\leftrightarrow$ x uniquely invented the zip).

And this clause does not give 'Julius' an object-dependent sense.

Thus, if we grant Evans's two background assumptions—that a referring expression is one whose contribution to truth conditions is stated by means of the relation of reference and that reference is just the relation that makes principle (P) come out true—then it is clear why 'Julius' is classified as a referring expression.[31] However, in order to understand why Evans rejects the idea that descriptive names belong semantically with definite descriptions, we need to see why definite descriptions cannot *also* be included in the category of referring expressions. So, why

[31] In his Preface to Evans's *The Varieties of Reference*, John McDowell says (pp. vi–vii): '[I]n notes for a lecture course on the theory of reference, Evans remarked that whereas some years previously he would have been tempted to call such a course "The Essence of Reference", now he would prefer to call it "The Varieties of Reference" . . . What he meant . . . was probably connected with his having become convinced that "descriptive names" are a perfectly good category of referring expressions. Earlier, he would have insisted that all genuine singular reference is . . . Russellian. Now that struck him as unwarrantedly essentialistic: a theoretically well founded conception of genuine singular terms could embrace both Russellian and non-Russellian varieties.'

is it that a statement of the semantic contribution of a definite description cannot be modelled on (16)?

The reason Evans gives is that such a statement of the semantic contribution of a definite description would not account for the way in which descriptions interact with modal operators. In possible-worlds semantics for modal languages, the satisfaction relation has to be relativized to worlds. So principle (P) must be replaced by (1979: 189):

(P′) If $R(t_1 \ldots t_n)$ is atomic, and $t_1 \ldots t_n$ are referring expressions, then $R(t_1 \ldots t_n)$ is true_w iff $<$ the referent of $t_1 \ldots$ the referent of $t_n >$ satisfies_w R.

But—and this is the crucial point—the relation of reference does *not* need to be relativized (1979: 189):

> Even in a modal language, all that is necessary to state the significance of names and other referring expressions is to state to what, if anything, they refer; the truth-with-respect-to-a-situation of a sentence containing a singular term depends simply upon whether or not its referent satisfies the predicate with respect to that situation. But, notoriously, this is not the case with definite descriptions.

It might be replied to this that there is something arbitrary about relativizing the relation of satisfaction but not the relation of reference. If we were to avoid this arbitrariness, and were to relativize the relation of reference to worlds, then definite descriptions could be grouped together with descriptive names and Russellian singular terms—ordinary proper names, indexicals, demonstratives—as referring expressions. But Evans's response to this proposal is that the use of a relativized relation of reference even for Russellian singular terms would involve an over-attribution of semantic powers. If we relativize the relation of reference in all cases then 'we ascribe to names, pronouns, and demonstratives semantical properties of a type which would allow them to get up to tricks they never in fact get up to' (1979: 190).

Evans's view, then, is that the decision not to relativize reference to worlds is well motivated rather than arbitrary. And if the relation of reference is not relativized, then descriptive names are grouped together with the familiar Russellian singular terms and are distinguished from definite descriptions. For descriptive names, like ordinary proper names, indexicals, and demonstratives, do not 'get up to tricks' in modal sentences. We do not, Evans says, use the descriptive name 'Julius' in such a way that sentences like:

> If you had invented the zip, you would have been Julius.
> If Julius had never invented the zip, he would not have been Julius.

come out true (p. 192; see also Evans 1982: 60).

A referring expression is one whose contribution to truth conditions is stated by means of a *non-world-relative* relation of reference that makes principle (P′) come out true. So, despite the modal and epistemic similarities between descriptive names and 'actually'-embedding definite descriptions, descriptive names do, and definite descriptions do not, belong in the semantic category of referring expressions.

7.2 Descriptive names in the two-dimensional framework

It may seem, however, that there is room for doubt as to whether Evans has really established that descriptive names are referring expressions, even by the lights of his own account of what a referring expression is. Davies and Humberstone raise this doubt by pointing out that, in a two-dimensional semantic theory for a modal language including '$\mathcal{F}$' as well as '$\Box$' and 'A' (1980: 12): 'The reference relation for proper names requires no relativization, that for descriptions requires the full double relativization, while the reference relation for descriptive names requires relativization in just the actual world place.' For clearly, there must be some world-relativity in the semantic axiom for a descriptive name such as 'Julius' in order to allow that the sentence:

(17) $\mathcal{F}\mathsf{A}$(Julius = Whitcomb L. Judson)

is false.

On the face of it, this doubt about Evans's argument turns on the behaviour of descriptive names within the scope of '$\mathcal{F}$', as in sentence (17). So Evans could respond to the doubt by returning to his reservations about the introduction of '$\mathcal{F}$' (Section 2). Certainly, if Evans is right to say that '$\mathcal{F}$' is a context-shifting operator, then there is a good reply for him to make. For, in that case, the relativization of the reference relation for descriptive names is nothing other than context-dependence, and even Russellian singular terms can be context-dependent. Thus, Evans says (this volume, 178):

> This naturally leads me to the disagreement I might have with you over the question of the need for relativizing the relation of reference to deal with 'Julius' in your '$\mathcal{F}$' contexts. I am quite happy to allow a relativity to a *context* [of utterance] is required once we accept as legitimate such [linguistic] contexts [in which 'Julius' occurs within the scope of '$\mathcal{F}$']. But I do not think that this marks a distinction between 'Julius' and other 'genuine' referring expressions since after all reference must be thus relativized for 'I', 'you', 'now' &c.

In fact, Evans's account of descriptive names as referring expressions could be defended without relying on the claim that '$\mathcal{F}$' is literally a context-shifting operator. Any objection to the introduction of '$\mathcal{F}$' would serve to defend the account against doubts that turn on the behaviour of descriptive names within the scope of '$\mathcal{F}$'. And an objection that extended to the introduction of '$\mathcal{D}$' for truth on the diagonal would defend the account against similar doubts that arise from the fact that the sentence:

(18) $\mathcal{D}$(Julius = Whitcomb L. Judson)

is false.

Suppose, for a moment, that there were good objections against the introduction of '$\mathcal{F}$' and of '$\mathcal{D}$', the operators that take advantage of variation in which world plays the role of the actual world. Then Evans's claim, that the contribution to truth conditions made by a descriptive name can be stated using a non-world-relative relation of reference, would be secure against doubts that depend on the properties of modal sentences such as (17) and (18). But there would still be other doubts that depend

on the modal properties of (non-modal) sentences. Thus, for example, we would still need to account for the fact that the sentence:

(3) If anyone uniquely invented the zip then Julius invented the zip.

is deeply necessary—true at every world considered as actual—even though there is a world in which Tiny Tim, rather than Whitcomb L. Judson, invented the zip. So the reference of the descriptive name 'Julius' must be allowed to vary as we consider the truth of (3) *in* different worlds. Similarly, the reference of 'Julius' must be world-relative in some way if we are to make sense of the idea that if a different world had been actual—if, for example, Tiny Tim rather than Whitcomb L. Judson had invented the zip—then 'Tiny Tim is Julius' would have been true.

Evans says (this volume, 179): 'I still cling to the idea that there is a *non-arbitrary* distinction which puts "Julius" with "Tom" [an ordinary proper name], and not with descriptions.' For the reasons just given, I think that descriptive names and ordinary proper names belong in different semantic categories. But it does not follow that descriptive names belong in the same semantic category as definite descriptions. Although Davies and Humberstone suggested that a descriptive name abbreviates an 'actually'-embedding description, they went on to say (1980: 11): 'Whether the suggestion ultimately proves to be tenable would depend on the resolution of such questions as: could a language containing unstructured expressions functioning as descriptive names fail to contain anything corresponding to "actually"?' Considerations of semantic structure might very well provide grounds for placing descriptive names in a *different* semantic category from definite descriptions.

It seems that we need a three-way distinction here. Ordinary proper names belong in a semantic category of Russellian singular terms. For members of this category, there is a semantic connection between the singular term and its referent and not just between the singular term and a descriptive condition. So there is no prospect of variation in reference without a change of meaning.

Definite descriptions belong in a different semantic category—arguably, in the category of quantifier expressions. In general, a definite description, 'The G', has a world-relative denotation because, as Evans says, the predicate 'G' has a world-relative satisfaction condition. Whether a given object satisfies 'G' varies as we move along the horizontal dimension of a two-dimensional array. When a definite description contains the 'Actually'-operator, this cancels out the horizontal world-relativity, but allows for variation in denotation as we vary which world plays the role of the actual world.

Descriptive names do not exhibit the horizontal world-relativity of definite descriptions, and they do not 'get up to tricks' when they occur within the scope of '$\Box$' or '$\Diamond$', or within the scope of modal operators in natural language. But they do still show some kind of world-relativity. For, as we have seen, the reference of a descriptive name varies (without any change in meaning) as we consider it *in* (but not *with respect to*) different possible worlds. This variation in reference can be conceived as resulting

from variation in which world plays the role of the actual world—variation as we move along the vertical dimension of a two-dimensional array. Thus, descriptive names, like 'actually'-embedding descriptions, exhibit vertical world-relativity. But, in the absence of horizontal world-relativity, moving along the vertical dimension comes to the same thing as moving along the diagonal. So we could equally well say that descriptive names and 'actually'-embedding descriptions exhibit diagonal world-relativity. And this way of putting it connects more directly with deep necessity and with truth *in* worlds.

Conclusion

My aim has been to plot the relationships between the notions of necessity that Humberstone and I characterized in terms of the operators '$\Box$' and '$\mathcal{F}\mathsf{A}$', Evans's notions of superficial and deep necessity, and the epistemic notion of *a priority*.

In the two-dimensional framework, the necessity expressed by '$\Box$' is truth on the horizontal, H-necessity, and the necessity expressed by '$\mathcal{F}\mathsf{A}$' is truth on the diagonal, D-necessity. Evans had reservations about the introduction of '$\mathcal{F}$', partly because of worries about utterance difficulties (Section 2). But, in any case, I have argued (Section 3) that Evans's superficial necessity is H-necessity, while his deep necessity coincides with D-necessity. Evans said that the combination of *a priority* with deep contingency would be intolerable and I have noted two problems about that claim. More importantly, I have suggested that the relationship between *a priority* and deep necessity may be asymmetric (Section 4).

Examples using descriptive names present no challenge to the coincidence of *a priority* with deep necessity, but examples using ordinary proper names threaten the inference from deep necessity to *a priority* (Section 5). A general strategy for maintaining the coincidence between *a priority* and deep necessity is to treat all referring expressions as being relevantly similar to descriptive names. But I have argued (Section 6) that this strategy faces objections similar to Kripke's objections to descriptivist theories of reference.

Finally (Section 7), I have expressed some reservations about Evans's own account of descriptive names, according to which they belong in a category of referring expressions alongside Russellian singular terms. However, neither Evans's account, nor my reservations about it, cast any doubt on the modal and epistemic similarities between descriptive names and 'actually'-embedding definite descriptions that are at the heart of Evans's solution to the puzzle of the contingent *a priori*.

References

Baldwin, T. (2001). On considering a possible world as actual. *Proceedings of the Aristotelian Society Supplementary Volume* 75, 157–74.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Burge, T. (1988). Individualism and self-knowledge. *Journal of Philosophy* 85, 649–63.

Burge, T. (1993). Content preservation. *Philosophical Review* 102, 457–88.
Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
Crossley, J. N. and Humberstone, I. L. (1977). The logic of 'actually'. *Reports on Mathematical Logic* 8, 11–29.
Davies, M. (1981). *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. London: Routledge & Kegan Paul.
——(1983). Actuality and context dependence II. *Analysis* 43, 128–33.
——and Humberstone, I. L. (1980). Two notions of necessity. *Philosophical Studies* 38, 1–30.
Donnellan, K. S. (1977). The contingent *a priori* and rigid designators. In P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy 2: Studies in the Philosophy of Language*. Minneapolis: University of Minnesota Press, 12–27.
Evans, G. (1979). Reference and contingency. *The Monist* 62, 161–89. Reprinted in *Collected Papers*. Oxford: Oxford University Press, 1985, 178–213. Page references to reprinting.
——(1982). *The Varieties of Reference*. Oxford: Oxford University Press.
——(1985). Does tense logic rest upon a mistake? In *Collected Papers*. Oxford: Oxford University Press, 343–63.
Field, H. (1996). The aprioricity of logic. *Proceedings of the Aristotelian Society* 96, 359–79.
Forbes, G. (1983). Actuality and context dependence I. *Analysis* 43, 123–8.
——(1989). *Languages of Possibility: An Essay in Philosophical Logic*. Oxford: Basil Blackwell.
Frege, G. (1892). Über Sinn und Bedeutung. Translated as 'On sense and reference', in P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell, 1952, 56–78.
Hazen, A. (1976). Expressive completeness in modal languages. *Journal of Philosophical Logic* 5, 25–46.
Humberstone, I. L. (2004). Two-dimensional adventures. *Philosophical Studies* 118, 17–65.
Jackson, F. C. (1998a). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
——(1998b). Reference and description revisited. In J. E. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind, and Ontology*. Malden, MA: Blackwell Publishers, 201–18.
——(2004). Why we need A-intensions. *Philosophical Studies* 118, 257–77.
Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, 481–563.
Kroon, F. (1987). Causal descriptivism. *Australasian Journal of Philosophy* 65, 1–17.
——(2004). A-intensions and communication. *Philosophical Studies* 118, 279–98.
Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
Lewis, D. (1980). Index, context, and content. In S. Kanger and S. Öhman (eds.), *Philosophy and Grammar*. Dordrecht: Reidel, 79–100. Reprinted in *Papers in Philosophical Logic*. Cambridge: Cambridge University Press, 1998, 21–44. Page references to reprinting.
Peacocke, C. (2004). *The Realm of Reason*. Oxford: Oxford University Press.
Pryor, J. (2000). The skeptic and the dogmatist. *Noûs* 34, 517–49.
Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.
Stalnaker R. C. (1978). Assertion. In P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*. Academic Press, 315–32. Reprinted in Stalnaker, *Context and Content*. Oxford: Oxford University Press, 1999, 78–95.

——(2001). On considering a possible world as actual. *Proceedings of the Aristotelian Society Supplementary Volume* 75, 141–56.
——(2003). Conceptual truth and metaphysical necessity. In *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*. Oxford: Oxford University Press, 201–15.
Taylor, B. (1976). States of affairs. In G. Evans and J. McDowell (eds.), *Truth and Meaning: Essays in Semantics*. Oxford: Oxford University Press, 263–84.
——(1985). *Modes of Occurrence*. Oxford: Basil Blackwell.

6

Comment on 'Two Notions of Necessity'

Gareth Evans

I confess to being a bit suspicious of the way you introduce your operator '$\mathcal{F}$', though I am quite unable to express my doubts in a compelling way.

Incidentally, I think the general ideas of your paper would be more clearly visible if you had taken as basic an operator '$\boxtimes$' with the condition:

$$W \models_w \boxtimes \alpha \text{ iff } (\forall W')(\forall w')[\text{if } W' \approx W \text{ then } W' \models_{w'} \alpha]$$

because this is closer to a necessity operator right from the start. But there are probably many refinements which would be difficult later, and it would not have the same degree of continuity with the earlier paper [Crossley and Humberstone, 1977].[1]

Anyway, definable in terms of your apparatus is an operator 'Poss' such that 'Poss(Actually(P))' is true provided '$\Diamond$P' is true (assuming P doesn't contain any descriptive names or 'Actually's). And this seems to me very like an operator in tense logic 'Φ(Now(P))' which is true provided '**F**(P)' is true.[2] Within the scope of 'Φ', 'Now' does not refer to the time of utterance; so equally within the scope of 'Poss', 'Actual' does not refer to the actual world. Yet in all other contexts 'Now'/'Actual' are intended to have the same role. (The kind of difficulty I am getting at will also emerge with 'Julius': within the scope of 'Poss', 'Julius' will not refer to the inventor of the zip.[3]) Now, I am able to make sense of these forms of embedding only if I understand them as involving a quite *new* form [of] embedding—quite unlike those previously recognized—of the kind I attempted to characterize under T_3 of my paper

Letter dated 14 July 1979, written to Martin Davies in response to a draft version of 'Two notions of necessity'. We are grateful to Antonia Phillips for her permission to publish this material. Notes have been added by MD.

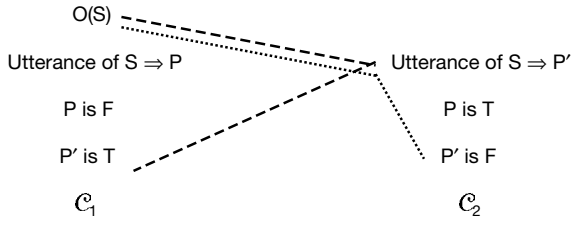
[1] Evans here proposes that we should take as primitive an operator equivalent to the combination '$\mathcal{F}\Box$'. An alternative proposal with a similar motivation would be to take as primitive an operator equivalent to the combination '$\mathcal{F}$**A**'.

[2] In his hand-written letter, Evans has, not 'Φ', but an inverted capital 'F'.

[3] In 'Reference and contingency' (1979/1985), Evans considers a descriptive name, 'Julius', introduced by the stipulation: 'Let us use "Julius" to refer to whoever invented the zip' (p. 181). If, as Davies and Humberstone (1980: 11) tentatively suggest, the descriptive name 'Julius' is semantically similar to the definite description 'the actual inventor of the zip', then the reference of 'Julius' will be world-relative when 'Julius' occurs within the scope of '$\mathcal{F}$' or 'Poss'.

on tense logic.[4] The semantic value of 'Poss(X)' upon an occasion of utterance, $\mathcal{C}$, is not a function of the semantic value X has upon that occasion, but the semantic value X *would* have upon some *other* (potential) occasion of utterance. Now I did not think and do not think that this form of embedding is incoherent, but I should like its distinctness from previously recognized forms to be made explicit.

[Actually, there may be a slight problem for understanding your 'Poss' in this way. Diagrammatically, and quantifying over propositions we have



T_3 in the tense case involves showing O(S) to be true in $\mathcal{C}_1$ because what S *would* have expressed in $\mathcal{C}_2$ *is* actually true (is true in $\mathcal{C}_1$) (this is marked out with the dashed line) whereas your 'Poss' involves, on this account, the route taken by the dotted line. And it may not be easy to prevent 'utterance difficulties' from getting in the way.][5]

[4] In 'Does tense logic rest upon a mistake?' (1985), Evans considers three conceptions of the semantic foundations of tense logic and, in particular, three interpretations of the temporally relative truth predicate 'true$_t$'. According to the third of these accounts, there is a direct connection between the truth$_t$ of a sentence S and the correctness of an utterance at t of S. Given such an interpretation of truth-at-a-time, we need to consider how to understand a recursive clause such as: For any time t, and any sentence S, true$_t$('P'^S) iff there is a time t', earlier than t, such that S is true$_{t'}$. Evans says (1985: 357): '[I]t is important to be clear about the novelty of this proposal, for it involves the recognition of a hitherto unknown form of embedding. In all previously-studied forms of embedding . . . the semantic value which a complex sentence $\Sigma(e)$ has in a given context is a function of the semantic value which the expression e has *in that context*. . . . But T_3 asserts that the semantic value which the sentence 'P(X)' has in a context is a function of the semantic value which X would have in *another* context. For, on the present interpretation, the recursive clause . . . says *roughly* that the utterance of 'P(X)' is true iff the utterance of X at some earlier time would have been true. If T_3 is right, the interpretation of a tensed utterance forces us to consider the interpretation which other, perhaps only potential, utterances would have, and this is a quite unprecedented feature.'

[5] The parenthetical worry that Evans raises here is related to the way that he improves on the 'rough' construal of the recursive clause quoted in the previous note. He explains the problem that arises with the rough construal as follows (1985: 358): 'Any such semantics must allow that the sentence 'In the past (There are no speakers)', as uttered now, expresses a truth, and to this end it must be the case that there is a time t', earlier than now, such that 'There are no speakers' is true$_{t'}$. But this cannot mean that, had someone uttered the sentence at t', he would have spoken correctly for he would not have done so.' He goes on to say that the problem—an example of what he calls 'utterance difficulties'—is 'a perfectly general one' and proposes a refined understanding of the relation between truth or correctness for utterances and truth-at-a-time for sentences (1985: 360): 'We want to speak of the *actual* value of a *potential* utterance; a sentence type is true$_t$ iff, were anyone to utter it at t, what he *would* thereby say *is* (as things stand) true.' In the case of the temporal operator, 'In the past', T_3 can avoid the utterance difficulty because we first consider the

Anyway, this is the only way I can understand 'Poss'—involving the thought of the utterance of the embedded sentence in other circumstances. So the question is: does it capture (thus understood) the notion of deep contingency. Is a sentence deeply contingent iff there is some possible circumstance in which its utterance would have produced a false* utterance? (qua sentence of the language) [* for our purposes this is all that matters] Well, provided the 'utterance difficulties' mentioned above can be dealt with, I think the answer is 'Yes'—but that is quite a big proviso since obviously on the most superficial reading 'I exist' would turn out to be deeply necessary.

I take it that you are opposed to this way of understanding 'Fixedly' &c. You would see a clear distinction (?) between 'Poss(Julius is Davies)' and 'To the left (I am hot)', and I am not sure that you are wrong.[6] But I would be grateful for a word or two more on this.

This naturally leads me to the disagreement I might have with you over the question of the need for relativizing the relation of reference to deal with 'Julius' in your '$\mathcal{F}$' contexts. I am quite happy to allow a relativity to a *context* is required once we accept as legitimate such contexts. But I do not think that this marks a distinction between 'Julius' and other 'genuine' referring expressions since after all reference must be thus relativized for 'I', 'you', 'now' &c. (I'm not sure how you would expect these context-dependent referring expressions to embed inside '$\mathcal{F}$'.) Perhaps I should have

possible situation in which there is an utterance at t of 'There are no speakers', and then consider the truth value as things actually stand, or in the actual world, of (what is said in) that possible, or potential, utterance at t.

But it is not so easy to use this strategy for avoiding the utterance difficulty when we consider the modal operator 'Poss' instead of the temporal operator 'In the past'. The sentence '$\Diamond$(There are no speakers)' is surely true. So the sentence 'Poss(Actually(There are no speakers))' should also be true. As Evans understands 'Poss', this must turn roughly on the truth of an utterance, u, of 'Actually(There are no speakers)' in some other possible situation, w. But here we run into the utterance difficulty, and it cannot be avoided by adopting the refinement of considering the truth value in the actual world, @, of what is said in the utterance, u, in w. There are two reasons for this. One reason is that following Evans's dashed line back to the actual world does not help. If what is said in u is that there are no speakers, this is no more true in @ than it is in w. The second, more general, reason is that the putative refinement gives the wrong truth conditions for other sentences. Thus, for example, the sentence '$\Diamond$(There are no tigers)' is surely true; so the sentence 'Poss(Actually(There are no tigers))' should also be true. As Evans understands 'Poss', this must turn roughly on the truth of an utterance of 'Actually(There are no tigers)' in some other possible situation. If what is said in such an utterance is that there are no tigers, then this may indeed be true in some possible situation; but it is certainly not true in the actual world. So it is the dotted line and not the dashed line that we need to follow.

[6] Evans here alludes to an example of the 'hitherto unknown form of embedding' that he provides in 'Does tense logic rest upon a mistake?' (1985: 357–8): 'Suppose that there is a language exactly like English, save that it possesses two additional operators, "To the right", and "To the left", which can be prefixed to sentences in the first person. A sentence like "To the left (I am hot)" as uttered by a speaker x at t is true iff there is at t on x's left someone moderately near who is hot.' Evans goes on to say that the only way to understand the construction as generating these truth conditions ('while continuing to suppose that the only role of the first person pronoun is that of denoting the speaker') is to suppose that the operator 'To the left' is governed by a rule: 'To the left'^S is true, as uttered by x at t iff there is someone moderately near to the left of x such that, if he were to utter the sentence [S] at t, what he would thereby say is true (p. 358).

said that a referring expression is any expression whose semantic contribution is dealt with by a relation of reference unrelativized save to deal with context-dependence (i.e. save to context). I still cling to the idea that there is a *non-arbitrary* distinction which puts 'Julius' with 'Tom', and not with descriptions.[7]

So you would expect me to dissent from your suggestion that a descriptive name is a conventional abbreviation for a definite description embedding 'actually'. I am impressed by the fact that we can introduce such names using the relation of reference

Let us refer to the ϕ by 'α'

and that this by itself guarantees rigidity in modal and temporal contexts for if we attempt to use 'α' non-rigidly, say by uttering

You might have been α

to mean (or on the basis of) 'You might have been the ϕ', we will be *infringing* the introducing stipulation, because we are certainly not there using 'α' to refer to the ϕ (or *as a name for* the ϕ). And this is why I am so resistant to regarding the second relativity in the truth relation as in any way similar to the first, and why I insist on regarding it as a form of context-dependence. Because this is the only way I can reconcile the truth of 'Poss(Tom = Julius)' with the stipulation that 'Julius' should be a name of the inventor of the zip. (Cf. the remark I made about 'To the left (I am hot)' being consistent with 'I' having the role of referring to the speaker.[8])

Incidentally I thought you might use your apparatus to say a word about Dummett's 'St. Anne must be a mother'.[9]

I thought the applications of the idea of a descriptive name and related ideas were fine. You raise a fascinating question about the difference 'acquaintance' plays in the case of water and of a particular spatio-temporal individual. I have thought and written about this, but it is all too long and probably too confused to put in a letter.[10]

One other minor point:[11] (p. 2) You write '$\mathcal{F}A\alpha$ says: whichever world had been actual, α would have been the case in the actual world'. But precisely because of the

[7] Evans here alludes to Davies and Humberstone (1980: 8), where the descriptive name 'Julius' is compared with an ordinary proper name, 'Tom', and with the description, 'the inventor of the zip'.

[8] See again the passage from Evans (1985: 358) mentioned in note 6.

[9] Dummett (1973: 113): 'After all, even though there is an intuitive sense in which it is quite correct to say, "St. Anne might never have become a parent", there is also an equally clear sense in which we may rightly say, "St. Anne cannot but have been a parent", provided always that this is understood as meaning that, if there was such a woman as St. Anne, then she can only have been a parent.'

[10] In his Preface to *The Varieties of Reference*, John McDowell says that Evans intended 'to reinforce the chapter on proper names with a partly parallel chapter on natural-kind terms' (p. vii). This material has never been published (but see the Index of Evans (1982), under 'Natural-kind terms and concepts').

[11] Evans actually wrote 'Two other minor points'; but the second of these is a genuinely minor typographical suggestion. The first point, in contrast, connects with one of the main themes of his letter. In order to avoid the problem that Evans raises here, the printed version of 'Two notions of necessity' has (Davies and Humberstone 1980, p. 3): 'Thus "$\mathcal{F}A\alpha$" says: whichever world had been actual, α would have been true at that world considered as actual.'

'rigidity' of 'actual' I hear this wrong; suggest you alter it to '. . . α would have been the case in that world'.

References

Crossley, J. N. and Humberstone, I. L. (1977). The logic of 'actually'. *Reports on Mathematical Logic*, 8, 11–29.

Davies, M. and Humberstone, I. L. (1980). Two notions of necessity. *Philosophical Studies*, 38, 1–30.

Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth.

Evans, G. (1979). Reference and contingency. *The Monist*, 62, 161–89. Reprinted in *Collected Papers*, 178–213 (page references to reprinting).

____(1982). *The Varieties of Reference*. Oxford: Oxford University Press.

____(1985). Does tense logic rest upon a mistake? In *Collected Papers*. Oxford: Oxford University Press, 343–63.

7

Two-Dimensionalism: A Neo-Fregean Interpretation

Manuel García-Carpintero

1. The Kripkean Puzzles

Saul Kripke's (1980) *Naming and Necessity* changed the assumptions defining the philosophical landscape of its times. A well-known case in point concerns Quine's presuppositions about quantificational modal logic. For Quine, the fact that use of this logical theory commits one to Aristotelian essentialism was enough to discredit any serious applications of it.[1] Unlike philosophers such as Carnap, Quine doubted that there was a distinctive class of necessary truths, but he shared with them the empiricist assumption that, if one exists, it coincides with the class of analytic and *a priori* truths: necessity has a linguistic foundation, if it has any at all, which for Carnap and other empiricists meant a foundation on convention.

Kripke proposed compelling examples and used them as a basis for providing clear-cut distinctions and forceful arguments. He distinguished between genuinely referential and descriptive denoting expressions. He argued that referential expressions like indexicals and demonstratives, proper names and natural kind terms are *de jure* rigid designators. This distinguishes them from other singular terms like definite descriptions, which might also behave *de facto* as rigid designators, but *de jure* are not so. By doing so he blunted the force of the only argument that Quine has provided against essentialism, based on the claim that no object instantiates *de re* essentially or contingently any property, but only relative to different ways of referring to it:

I would like to thank Jose Díez, Dan López de Sa, Josep Macià, Manuel Pérez and David Pineda for helpful comments, and Michael Maudsley for his grammatical revision. This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the Spanish Government's grant DGI BFF2002-10164 and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409, from DGI HUM2004-05609-C02-01, DURSI, Generalitat de Catalunya, SGR01-0018, and a *Distinció de Recerca de la Generalitat, Investigadors Reconeguts* 2002–2008.

[1] According to Quine the commitment to Aristotelian essentialism does not lie in the fact that a proposition stating it is a theorem of the logical theory, but is rather of a pragmatic nature. See Burgess (1998) and Pérez-Otero and García-Carpintero (1998).

even if the world's tallest mathematician is in fact the world's tallest cyclist, he is not *de re* necessarily rational or two-legged, but only *de dicto*, necessarily rational as the world's tallest mathematician, necessarily two-legged as the world's tallest cyclist. For this Quinean argument crucially depends on overlooking the distinction between *de jure* rigid designators and other designators. Relatedly, and also importantly, Kripke distinguished what we might call epistemic necessity from metaphysical necessity. Some truths, he argued, are *a priori*, but nonetheless contingent; some other truths are necessary, but nonetheless *a posteriori*.[2]

Although in this way Kripke undermined dogmatic rejections of essentialism based more on philosophical prejudice than on sound argument, he was nonetheless well aware of the main philosophical puzzle created by his proposals, which, beyond philosophical dogma, probably accounts for the traditional identification of the modalities throughout the history of the discipline. As rehearsed in the introduction to this volume, Kripkean views about referential expressions envisage *modal illusions*: truths that are in fact necessary appear to be contingent, including instances of the schema ***if n exists, n is F***, with a rigid designator in the place of 'n' and a predicate signifying a hidden essential property of its referent in the place of 'F'. To use the standard illustration, let us replace 'F' in the schema with 'is-identical-to-Hesperus' and 'n' with 'Phosphorus':

(1) If Phosphorus exists, Phosphorus is-identical-to-Hesperus

The existence of those modal illusions elicited by Kripke's compelling views about the metaphysics of modality is puzzling in view of the plausibility of another influential view that Kripke espoused, concerning this time the epistemology of modality: that a possible world "isn't a distant country that we are . . . viewing through a telescope . . . 'Possible worlds' are *stipulated*, not *discovered* by powerful telescopes" (Kripke 1980, 44); "things aren't 'found out' about a counterfactual situation, they are stipulated" (1980, 49).[3] This suggests that we have reasonably reliable *a priori* access to possible worlds. For Kripke's remarks are made in the context of a criticism of Lewis's view that possible worlds are concrete, like the actual world of which we are part. His remarks thus appear to implicitly advocate alternative views such as Stalnaker's that possible worlds are properties that the only actual world might have. Views like these are typically defended on the basis that they provide for a more sensible epistemology of modality, allowing that we in general know the modal facts that we take ourselves to know.

Kripke's stated remarks are characteristically cautious; my own rendering, as the explicit claim that we have a reasonably reliable *a priori* access to modal facts, is much less so. But this interpretation appears to be closer to the text than, for instance, Soames' (2003, 356) deflationary reading, according to which Kripke is just saying in those passages that it is up to us to stipulate, or specify, which of the possible states that the world genuinely could have been in that we are interested in, and wish

[2] Soames (2003, 347–54) provides an excellent presentation of these issues.

[3] See also the analogous remarks in Kripke's (1980) preface that possible worlds are "given" by descriptive stipulations, pp. 15–20.

to make claims about. This cannot be Kripke's point, because, as I have said, his claims are intended as an alternative to Lewis's realism about possible worlds; and the realist about possible worlds will also accept that we stipulate situations in Soames' understanding of *stipulation*.

This Kripkean puzzle does not merely arise in some isolated cases; on the contrary, a systematic pattern is predicted. To defend the core Kripkean views about *de re* modality requires thus a philosophical account of these systematically predicted illusions, consistent with modal knowledge. The promise of doing this is precisely what, in my view, makes 2-D accounts appealing. Here I would like to explore a neo-Fregean elaboration of the 2-D central idea that (in Kripke's terms) "an appropriate corresponding qualitative statement", different from the original, necessary one, which unlike this "might have been false", is somehow mixed up with it, thus engendering the illusion of its contingency. What makes the proposal neo-Fregean is that, instead of assuming an independently given epistemic conception of the modal realm on which primary intensions are built, the account derives it from a Fregean-like distinction between the sense and the reference of the expressions systematically responsible for the Kripkean cases.

On the view to be presented here, the availability of the core 2-D model for the necessary *a posteriori* is dependent on the account also applying to the other puzzling Kripkean category of the contingent *a priori*. To go back to the earlier examples given in the introduction to this volume, as Kripke notes, if one stipulates that a designator *N* is to be used to refer to an object introduced by a description *D* that thus fixes its reference, one can be said to know thereby *a priori* "in some sense" (1980, 63) the truth of the corresponding statement '*N* is *D* if *N* exists'; (2) provides the conventional example related to (1):

(2) Phosphorus is whatever appears as shining brightly in the east just before sunrise, if it exists.

Here the model should explain how, although (2) signifies a contingent proposition, there is "an appropriate corresponding qualitative statement" which expresses a necessary one.

I will present my proposal as an elaboration of Stalnaker's (1978) 2-D framework—unfaithful to some of Stalnaker's crucial assumptions. My strategy will be to critically examine Stalnaker's recent scepticism about its explanatory credentials. Considering just two possible states of the world, the following matrix was used in the introduction to represent the proposition expressed by (2); *i* is the actual state of the world, and *j* an alternative state relative to which it is Mars that appears as shining brightly in the east just before sunrise, otherwise as close as possible to the actual state of the world:

A

i	*j*
T	F

Worlds *i* and *j* were also used to illustrate the second way in which the truth-value of what we utter depends on the facts, emphasized by Stalnaker: if the facts had been different, what one says might have differed too. Given the astronomical facts as they are relative to *j*, if the stipulation fixing the reference of 'Phosphorus' in *i* still prevails in *j*, (2) expresses a different proposition, one about Mars; we can represent this second way in which the truth-value of what is expressed is determined by the facts by adding a second row:

B

	i	*j*
i	T	F
j	F	T

Now, this propositional concept for (2) includes a necessary diagonal proposition. Of course, this is so only because we have kept fixed an aspect of the facts determining the different contents that the very same utterance might have had, the reference-fixing description tied to 'Phosphorus'; but there is nothing in Stalnaker's metasemantic conception of a diagonal proposition as such requiring it. Taking into account a possible world *k* in which it is stipulated that 'Phosphorus' refers to the innermost planet in the Solar System, otherwise as close to the actual world as it could be, in particular such that in *k* Venus is the brightest heavenly body seen in the morning, we get:

C

	i	*j*	*k*
i	T	F	T
j	F	T	F
k	F	F	F

This poses a problem for 2-D alleged accounts of the contingent *a priori*, to which I will come back presently. But at this point I need to examine carefully Stalnaker's Gricean suggestions to account for the necessary *a posteriori*. The official Kripkean content of (1) is a necessary proposition with the following partial matrix:

D

i	*j*
T	T

Given the facts about *j* indicated above, a corresponding propositional concept would be partially represented by the following matrix:[4]

E

	i	*j*
i	T	T
j	F	F

This propositional concept does not offer a clear indication of how to modify the context set. According to E, what the speakers should do is: if the actual world is *i*, then keep both *i* and *j* in the context set; if it is *j*, then eliminate both *i* and *j* from the context set. Since the speakers do not know which of *i* or *j* is actual, they do not know how to proceed on the basis of E, and so asserting (1) would have no significant effect on the context set. This is why, on the basis of Gricean considerations, speakers are assumed to infer that the content asserted is not one of the horizontal propositions, but the diagonal proposition: "in these special cases, the horizontal propositions of the propositional concept do not themselves represent what is said: they represent what is said *according to the normal semantic rules* as they are in the relevant possible world. In such a case, the normal semantic rules are overridden"; rather, in cases such as these "one should identify what is said with the diagonal proposition of the propositional concept determined by the context" (Stalnaker 1999, 13–14).

The question I want to press now is this: what reason is there, given Stalnaker's assumptions, to include the row corresponding to world *j* in the representation of such a context? What reason is there, in other words, to assume that it is compatible with the common ground that 'Hesperus' and 'Phosphorus' have meanings such that (1) might convey a (necessarily) false proposition, like the one partially characterized in E for *j*? For, without such a justification, the Gricean argument does not get started.

A natural response is as follows: "In a context where the hearer knows the full meanings of the terms used in an utterance (for example, if they know that 'Hesperus' and 'Phosphorus' both refer to Venus), and where this knowledge is common ground between speaker and hearer, then the utterance will convey its original propositional content. But if the hearer does not know the meanings of the terms, then the utterance will convey a different content" (Chalmers (forthcoming), section 2.2 on Stalnaker's views). But this cannot be a good reason in general, still less for someone with Millian views like Stalnaker. According to those views, the meaning of a proper name is simply its referent. To include worlds like *j* in the context set, the argument

[4] This is not strictly speaking correct. Worlds *i* and *j* in *A* and *B* should be taken as centred around transworld counterparts of the relevant utterance of (2); while in *C* and *E* they must be centred around transworld counterparts of the relevant utterance of (1).

appeals to lack of knowledge of meaning, assuming that this will occur whenever the speaker does not know that the two names refer to the same entity. However, there can be informative true identity statements for any name; because of this, (1) is just a convenient illustration of a well-known pattern: as I have emphasized, there are systematic ways of producing statements creating the Kripkean modal illusions. Hence, the response we are considering in fact requires that speakers never know the meaning of the names they use, no matter how well acquainted they are with their referents. Given any proper name that a speaker uses, we can always produce examples such as (1) involving it. To deploy the Stalnakerian proposal, we would need propositional concepts including several rows, such as *i* and *j* in E. In order to appeal to the present justification, we should assume that speakers never know the meaning of the name.[5]

What Stalnaker says is not very helpful on this matter. A reason he provides to assume propositional concepts like E in these cases is simply that the diagonal proposition then obtained through the Gricean consideration is "an intuitively plausible candidate for the information that speakers intend to convey in such contexts" (1999, 13); "it is clear that the diagonal proposition is the one that the speaker means to communicate" (1978, 92). I take this to be so, but what is at stake is whether Stalnaker's theoretical assumptions are compatible with this desired result. A second reason he provides is that "to construct a context . . . in which the proposition expressed is neither trivial nor assumed false, one must include possible worlds in which the sentence, interpreted in the standard way, expresses different propositions" (1978, 92). But this appears to be as question-begging as the previous point; for we know we must have propositional concepts with the structure of E to avoid the result that the proposition expressed is either trivial or assumed false, but the issue is whether our theoretical assumptions allow us to avoid it. My first concern in what follows is to provide a theoretical proposal that has the desired effects. As announced, it is part of a neo-Fregean view of reference, far removed from Stalnaker's sympathies; I will not discuss any longer whether other theoretical accounts more accommodating of his views would allow for similar results, although I very much doubt it.

2. Stalnaker's Challenge

Let us now go back to Stalnaker's worries about 2-D accounts of the contingent *a priori* posed by the need to exclude worlds like *k* in C above, which he presents relative to his metasemantic account with his "7 + 5 = 12" example:

> Consider a context in which a person is uncertain about whether the intended meaning of a certain token of "7 + 5 = 12" is the usual one, or one that uses a base 8 notation, with the same numerals for one through seven. In some possible worlds compatible with the beliefs of this person, the token expresses the falsehood that seven plus five is ten, and so the diagonal will be contingent. [. . .] So the metasemantic interpretation yields no account

[5] For a different twist to this worry, see Soames (2005, 96–9).

or representation of *a priori* truth or knowledge, and does not depend on any notion of the *a priori*. (''Assertion revisited'', this volume, 302–3.)

Stalnaker considers the objection that, in determining the epistemological status of a statement, possible worlds like the one he envisages for '$7 + 5 = 12$' in the quoted text are irrelevant, on the basis of what I will henceforth call (for reasons to be explained presently) *the meaning-constitution point* that Stalnaker describes as ''the familiar point that the necessity and *a priori* city of mathematical truths such as '$7 + 5 = 12$' is not compromised by the undisputed fact that it is only contingently true (and known only *a posteriori*) that we use arithmetical notation as we do'' (1999, 16).[6] However, as he reminds us, ''the two-dimensional apparatus was introduced for the purpose of representing (on the vertical dimension) variations in the propositions expressed'' (1999, 16), and this poses a challenge for anybody wanting to defend the view that worlds like *k* in C are irrelevant when determining the epistemological status of a statement: ''If we are to represent *a priori* truth by the necessity of the diagonal, we must either find grounds for excluding worlds [like *k*], or else find a different way of associating propositional concepts with utterance events'' (1999, 16). We have seen before what, it will turn out, in fact is a related concern with respect to 2-D accounts of *a posteriori* necessities like (1); namely, that we lack a justification to posit propositional concepts like E.

In what follows, I will be confronting this challenge. Let me henceforth call the singular propositions constituting what Stalnaker described as ''what is said *according to the normal semantic rules*'' by utterances of sentences such as (1) and (2) their *official contents*. The approach I will pursue will be to isolate an *a priori* component in the knowledge constituting understanding of the rigid designators contributing the relevant *res* to official contents. Diagonal propositions, I will suggest, model that *a priori* component. The concept of *apriority* thus modelled is one along the lines envisaged by Reichenbach (1920), according to which apriority in the relevant cases is a form of analyticity.

Contemporary epistemologists like Bealer (1999), Bonjour (1998), Burge (1993), or Peacocke (1993) have emphasized that sensible accounts of the *a priori* should be *moderate*, in allowing for the fallibility and defeasibility of what is taken to be *a priori* knowledge. In a defence of apriority in the face of the scientific rejection of Euclidean geometry, Reichenbach (1920) urged the severance of two elements in the Kantian conception of the *a priori, necessary and unrevisable truth, fixed for all time*, on the one hand, and *truth constitutive of the object of [a posteriori] knowledge*, on the other, arguing that only the former should be abandoned: '' '*A priori*' means '*before* [*a posteriori*] knowledge,' but not 'for all time' and not 'independent of experience''' (1920, 105). The elucidation of the 2-D framework I would like to suggest in what follows agrees with him on both counts.

[6] Kripke is reported to have made this point in the John Locke lectures on Reference and Existence: ''One should not identify what people would have been able to say in hypothetical circumstances, if they had obtained, or what they would have said had the circumstances obtained, with what we can say of these circumstances, perhaps knowing that they don't obtain.'' (This quote might not accurately represent Kripke's views.)

We have been discussing so far only examples involving proper names; it will help to consider related examples involving demonstratives, before going back to them:

(3) If that morning heavenly body exists, that morning heavenly body is-identical-to-that-evening-heavenly-body[7]

There are many controversial issues involving the semantics of indexicals in general and (complex) demonstratives in particular, but we do not need to go into them here; for present purposes, I will just take for granted an at least plausible position on some of them. It is a slight variation on Kaplan's (1989) views, strongly influenced by John Perry's work (1997, and references there), which I have defended in previous papers.[8]

To account for intuitions analogous to those motivating the Kripkean views about proper names, the view takes indexicals and demonstratives to be rigid designators; it is, of course, only contextualized, token-expressions that can be counted as designators at all in these cases, and thus references to linguistic items are to be henceforth understood as references to tokens. The contribution of a complex demonstrative like 'that morning heavenly body' in an utterance of (3) to the asserted content is according to the present view the same as the contribution of 'Phosphorus' in (1): the object referred-to.[9] This distinguishes the complex demonstrative from the similar description 'the morning heavenly body'; although the latter might be referentially used *de facto* as a rigid designator, *de iure* its contribution is quantificational.[10] On this view, matrix *E* should provide a partial representation of the propositional concept corresponding to (3) as good as it is for (1).[11]

It is however clear that in this case there is descriptive information concerning the referent of the complex demonstrative that any competent speaker would obtain from the utterance of (3). On the view I am outlining, this information is not part of the asserted content, but belongs in a different proposition, which is not asserted but presupposed, a conventional implicature.[12] Stalnaker's primary notion of *presupposition* is that of an attitude of speakers in particular contexts. Nevertheless, he acknowledges that, like meaning, referring, asking, implying and so on, presupposing is something that both speakers and the words they use can be said to do, and he contemplates thereby a notion of pragmatic *sentence* presupposition, a presuppositional *requirement:* "Sentence *S* presupposes that *P* if and only if the use of *S* would be inappropriate in a context in which the speaker was not presupposing that *P*" (1999, 7). He had already made it clear in "Assertion" that "the context on which an assertion has its *essential* effect is not defined by what is presupposed before the

[7] We are supposed to imagine (3) uttered with the factually required time-lag.

[8] See García-Carpintero (1998 and 2000).

[9] In the following discussion, I will ignore presuppositional effects created by the use of 'exists' in (3) and related utterances.

[10] Note, however, that some writers, including King (2001), have argued for an alternative quantificational account of complex demonstratives. I cannot properly go into this here.

[11] For reasons already given (see fn. 4), this must be taken *cum grano salis*.

[12] Dever (2001) defends a multi-propositional view of complex demonstratives, on which the descriptive proposition is not presupposed, but plays a different logical role. My view is also close in relevant respects to Glanzberg & Siegel's (forthcoming).

speaker begins to speak, but will include any information which the speaker assumes his audience can infer from the performance of the speech act" (1999, 86); this of course applies to the present case.

In a nutshell, the proposal is that the relevant presupposition corresponding to the complex demonstrative in (3) is, if we make it explicit, that *that token of 'that morning heavenly body' refers to whatever morning heavenly body is most salient when it is uttered* (where "that token of 'that morning heavenly body' " is intended to refer to the token of that expression in the relevant utterance of (3)).[13] If this proposal is correct, it provides a response to Stalnaker's challenge in the quotation at the beginning of this section. Applied to the present case, the challenge ultimately asks us to justify that a corresponding alleged instance of the contingent *a priori*, an utterance of (4) in (*mutatis mutandis*) the envisaged context for (3) expresses a necessary diagonal proposition (analogous to the one represented by *B* for (2)):

(4) That morning heavenly body is whatever morning heavenly body is most salient when that very token of 'that morning heavenly body' is uttered, if it exists.

For (4) to have a necessarily true diagonal, in each world in the envisaged context for (3) and (4) "that morning heavenly body" refers to the most salient heavenly body visible in the morning in that world; and so worlds (analogous to k in C) where, say, "that morning heavenly body" refers to Vincent van Gogh's left ear are excluded.

According to the view I am outlining, the *official* contribution to contents of a complex demonstrative is, like that of a proper name, a typically extra-linguistic and extra-mental object. These objects might well have hidden essential properties, which they will keep in every possible world compatible with those contents; as in the case of proper names, (3) ascribes to the demonstrative's referent a hidden essential property of this kind, *being identical to that evening heavenly body*, and this is why it expresses a necessary singular proposition. On the other hand, the properties used to fix reference to those objects might well be contingent, and this is how the official content (the horizontal) of (4) will be a contingent proposition.

But what reasons do we have to count the diagonal proposition for (4) as necessary, in view of Stalnaker's challenge? Is it not obvious that there are possible worlds in which the relevant utterance of (4) is made relative to a language in which the determiner 'that' in the complex demonstrative is interpreted in the way that we interpret 'every', everything else (including the meanings of the other expressions in the utterance) being as close as possible to actuality? Should we not consider worlds like these as part of the context, even if, relative to this sort of world, what the utterance of (4) says is false, and so a world like this has the same effect on the modal status of the diagonal as *k* in *C*? The view I am advancing is that, in this particular case, we can appeal to a surrogate of the *meaning-constitution point* that for rigid

[13] It must be assumed that speakers competent by ordinary standards are able to somehow grasp this implicitly, without having explicitly articulated concepts of, say, *reference, salience*, or the type–token distinction. This raises additional concerns about the present proposal for interpreting the 2-D framework that I am not in a position to address here.

designators Kripke put as follows: "When I say that a designator is rigid, and designates the same thing in all possible worlds, I mean that, as used in *our* language, it stands for that thing, when *we* talk about counterfactual situations. I don't mean, of course, that there mightn't be counterfactual situations in which in the other possible world people actually spoke a different language" (Kripke 1980, 77).

The main idea for the 2-D treatment of those examples of apriority that the Kripkean discussion highlights was, according to Stalnaker, that an "*a priori* truth is a statement that, while perhaps not expressing a necessary proposition, expresses a truth in every context" ("Assertion", 83). If this is so, in considering a possible world as actual, in order to determine the modal status of the relevant diagonal proposition, we should still be considering only the different propositions that expressions as used in *our* language could have meant. We should allow for variations in the referent of the complex demonstrative; but a situation like the previous counterpart of *k*, in which the complex demonstrative is not used at all as the demonstrative 'that', is not one in which the expression belongs in our language. Variations in the contribution that the complex demonstrative, as used in the actual world, makes to the asserted official content are allowed; variations in the descriptive condition that the referent is presupposed to satisfy, semantically derived from the linguistic meaning of the constituent NP and the simple demonstrative, are not.

The question is, of course, whether there is any justification for this invidious treatment of different semantic properties that the complex demonstrative has, as used in our language, in addition to our desire to ensure that (4) eventually expresses a necessary diagonal proposition. What I will be arguing in the following section is that the association of the complex demonstrative with a descriptive condition is constitutive of its meaning in a way that its association with its referent is not. This is why I am referring to Stalnaker's "familiar point" as the *meaning-constitution* point. I will defend that, in addition to providing a plausible justification that utterances like (4) express a necessary diagonal, the proposal gives an acceptable account of the nature of the *a priori* knowledge that the diagonal models. Last, but not least, the proposal will offer an immediate justification for considering propositional concepts with the structure of E for utterances of sentences like (1) and (3), thus allowing for the Stalnakerian account that the diagonal proposition, not the official content, is expressed in those cases, for which we could not find any proper rationale in Stalnaker's texts.

3. Constitutive Properties of Referential Expressions

Competent speakers will in principle be able to understand the official singular contingent proposition asserted in uttering (4). This understanding constitutes a piece of knowledge, and thus a justified belief. However, it is not merely speakers' linguistic competence that is involved in the justification of beliefs such as the one about the singular content signified by (4).[14] My main argument for this has two parts. First

[14] This claim coincides to a good extent, I think, with Soames' (2003, ch. 16) main point.

(A), in the particular case of the utterance of (4) that I am considering, the concurrence of a veridical perceptual experience of Venus will also be a substantive feature of that justification, well beyond what linguistic competence provides. Secondly (B), and more in general, the existence of the sort of relation with objects that those cases illustrate is in general a substantive part of the justification of every particular act of understanding singular contents about material objects like those we are considering, which similarly transcends linguistic competence.[15]

The 2-D treatment of cases of the contingent *a priori* such as (2) and (4) for which I will be arguing is in fact close to Donnellan's (1979). He claims that knowledge constituting understanding of singular propositions, like the official contents of (1)–(4), cannot be *a priori*; according to him, therefore, the Kripkean *a priori* knowledge of the truth of (2) and (4) cannot have those official contents as its objects. He makes a case for this in part by characterizing metalinguistic contents that, he suggests, more plausibly play the role of contents that are known *a priori*. I agree with Jeshion's (2001) objections to this part of his argument; she argues that the examples in which it is clear that speakers merely grasp metalinguistic information are manifestly unlike the ones that concern us, while in the case of analogous examples it is unclear that speakers grasp merely metalinguistic information. However, I think that the two-dimensional candidates for the relevant contents improve on Donnellan's, and can withstand the corresponding objections.

Like Jeshion, I also like a further aspect of Donnellan's discussion, namely, that it does not rely on the assumption that understanding singular contents always *requires* the presence of a non-conceptual relation of acquaintance (through perception, as in the examples so far, memory or testimony) with the singular elements. I would like to allow, with Donnellan and Jeshion, that competent speakers may grasp the *prima facie* singular contents expressed by utterances of sentences like (6), where the reference of the demonstrative is fixed relative to descriptive information provided by the previous discourse, (5) here:

(5) There is a single planet causing perturbations in Uranus's orbit
(6) That planet causes perturbations in Uranus's orbit, if it exists

Therefore, I cannot appeal in general merely to considerations such as A above for the substantive, beyond-the-linguistic character of competent understanding of the official singular contents expressed by utterances like (1)–(4); for point A is that (perceptual) acquaintance is part of what is required in order to properly understand them, but I am agreeing that this does not apply in cases like (6). Jeshion's (2002a) proposals improve our still poor understanding of what it is to grasp singular contents; according to her, what makes an attitude singular is not necessarily acquaintance, but its role in cognition; "What distinguishes *de re* thought is its structural or organizational role in thought" (2002a, 67).

[15] I say 'substantive' instead of 'empirical' because I want to allow for singular attitudes about abstract objects, like numbers and fictional characters, and, although I will not elaborate on this here, I want my considerations to apply *mutatis mutandis* to them.

Nevertheless, her account grants that there is something correct in views requiring acquaintance, in that they at least characterize the paradigm cases of *de re* contents: "Although I have argued that acquaintance is not necessary for *de re* belief, I have not argued that acquaintance is not in some way significant to an understanding of *de re* belief. *De re* beliefs via acquaintance *are* developmentally primary. Also, I would hypothesize that acquaintanceless *de re* belief is impossible without *de re* belief with acquaintance. And, no doubt, it is (direct) acquaintance that suggests the idea of a belief being directly about an object" (2002a, 70). My more general consideration B for the substantive nature of understanding singular contents will take its lead from this concession.

Singular contents, I am assuming, are *object-individuated*: different objects determine different singular contents. Some writers, notoriously including Evans and McDowell, take singular contents to be also *object-dependent*. Many philosophers find this view, as apparently understood by Evans and McDowell, unnecessarily strong. (5) and (6) are of course the equivalents involving complex demonstratives of similar cases concerning the proper name 'Neptune', under the usual assumptions about Leverrier's descriptive fixation of its reference. If we substitute 'Mercury' for 'Uranus' in them, we get similarly corresponding cases involving complex demonstratives of notorious actual examples of reference-failure with the proper name 'Vulcan', (7) and (8) below. Many writers find understandably implausible the view that, after the replacement, we move from utterances expressing official singular contents to utterances that do not express such contents.

(7) There is a single planet causing perturbations in Mercury's orbit
(8) That planet causes perturbations in Mercury's orbit, if it exists

Now, *dependence* can be explained in terms of essence; in those terms, a natural understanding of object-dependence, compatible with what Evans and McDowell assume in putting forward their views, is that the object(s) a singular content is about is (are) part of its constitutive essence.[16] There is, however, a weaker notion of dependence, which provides for a more plausible view, consistent with Jeshion's remarks on the dependence of acquaintanceless singular attitudes on acquaintance-based ones. On this view, while no actual relation to a particular object is part of the essence of particular singular contents, it is nevertheless part of their constitutive nature that they belong in a class of contents, some of which do involve acquaintance relations. Such a weaker notion of object-dependence for singular contents would allow cases of failure of reference like (8) to signify them.

This weaker notion of object-dependence will help to sustain an already familiar line of resistance to the McKinsey-style reasoning purporting to show the incompatibility of externalism and self-knowledge, clearly articulated by McLaughlin and Tye (1998, 367–71). I can know in a privileged way the thought that I am expressing when I put forward, say, (4). This is an object-dependent thought, in that it *aims* to be object-individuated. There are successful and unsuccessful varieties

[16] See Fine (1995) for the relation between *dependence* and *essence*, and for more on the distinction between *generic* and *specific* dependence that I am about to appeal to.

of such object-dependent thoughts. Given that the thought I am entertaining is successfully object-dependent, it follows from philosophical considerations that that heavenly body exists; and I am in a position to appreciate that this is the case. On the other hand, only empirical methods can justify my thought that that heavenly body exists. But there is nothing problematic in this package of thoughts; for it is only empirical methods that can justify my thinking that my thought belongs in the successful class. No amount of pure reflection and philosophical reasoning can achieve that feat.

I am now in a position to elaborate on part (A) of the argument, that is, that in the case of the utterance of (4) being considered, the concurrence of a veridical perceptual experience of Venus will also be a substantive feature of that justification, well beyond linguistic competence. Contemporary writers on the *a priori*, particularly Burge (1993), have made us sensitive to the distinction between perception (or other empirical justificatory methods) playing a merely *enabling* role, versus its playing a substantive justificatory role.[17] The distinction is subtle, and of no clear application in many cases (which is why ultimately I prefer the more positive Reichenbachian characterization of the *a priori* here taken to be articulated by the 2-D framework, to the more negative traditional one as non-empirical justification). Burge takes a justification to be *a priori* just in case "its justificational force is in no way constituted or enhanced by reference to or reliance on the specifics of some range of sense experiences or perceptual beliefs" (1993, 458). Is the justificational force of my justification for grasping the object-dependent thought expressed by, say, (4) so enhanced?

Now, consider: the thought I am thinking when I entertain (4) is different whether or not it is successfully object-dependent, for object-dependent thoughts are object-individuated. Whether or not it is successful crucially depends (in the present case) on whether or not I do actually perceive a heavenly body, as opposed to merely having some perceptual experiences. When I take the thought I am entertaining at face value, I assume it to be of the successful variety; this is part of what is meant by the idea that object-dependent thoughts, even in our weak characterization, *aim* at objects. Now, suppose that the relevant perception merely plays an enabling role, as opposed to a justificatory one, in this assumption of mine that I am entertaining a thought of the successful variety. In that case, I do not think we can stop the McKinsey-style derivation of my privileged, almost-*a priori* access to the claim that that heavenly body exists, and we should conclude, incorrectly I assume, that I am justified in thinking that it exists merely by a combination of reflection and philosophical methods.

Let us now move to part B of the argument, for the substantive character of understanding in acquaintanceless cases. Singular contents can be grasped in the absence of acquaintance with the relevant objects, as in (5)–(8). However, assuming Jeshion's concession, those cases only exist against the background of others that do involve acquaintance. Now, in cases involving acquaintance, it is as we have seen a crucial feature of the justification constitutive of our understanding singular contents that our evidence (the relevant perceptual experiences, in cases (1)–(4)) does put us *en rapport*

[17] For present purposes, I will not distinguish between *justification* and *entitlement*.

with objects; and this is—I have argued—a substantive, indeed empirical element going beyond mere linguistic competence: this was part A of the argument. Hence, in pure descriptive cases like (5)–(8) it is part of our *total evidence* justifying understanding the one acquired through empirical justification in cases involving acquaintance. I take this to be a similarly substantive justifying assumption. That this empirical collateral information plays a justifying role, and not merely an enabling one, can be established on the basis of the very same considerations developed in the previous paragraphs for the acquaintance cases. Acquaintanceless cases presuppose acquaintance cases; the total evidence constituting the justification for understanding in the former cases includes the one supporting cases of the latter variety.

I appreciate that these are relatively abstract considerations; to fill them up in sufficiently convincing detail, however, would require a better grasp of the nature of *de re* attitudes than I am in a position to provide here. I will try to make up for this with a few brief impressionistic remarks. On the present view, the official content *asserted* by means of (6) is a singular proposition, as much as it is in the case of (1)–(4); the descriptive material that my view also posits is part of a *presupposed* content. This fits the facts regarding our intuitions about their possible world truth-conditions that defenders of singular propositions have emphasized, to wit, intuitions indicating that it is how things are with the objects themselves, whether or not they fit the descriptive material, that is relevant for the truth of what is said relative to different possible circumstances. It also fits the facts regarding the propriety of *de re* reports of the relevant asserted contents, reports that satisfy the two well-known Quinean criteria of openness to correct applications of the logical laws of sustitutivity and existential generalization. Although the distinction between *de re* reports and *de re* attitudes—our true present subject—will never be sufficiently emphasized, there certainly must be some weak connection between the latter and the former of the kind suggested here.

What is it that those two sets of intuitions point to? Whenever he tries to characterize *de re* attitudes, Evans (1982, 146) offers suggestions such as this: "a subject who has a demonstrative Idea of an object has an *unmediated* disposition to treat information from that object as germane to the truth or falsity of thoughts involving that Idea." Imagine the following case. I am visiting an exhibition in a medieval cloister; I am carrying a heavy bag, and to unburden myself during the visit I put it inside a big porcelain vase in a corner. In fact there are perceptually indistinguishable vases of this kind in each corner in the cloister. Some time later, I judge, in front of one such vase: *this vase contains my bag*. It may well be that nothing of a purely general character that I can have access to in my full conscious state (nothing *descriptive*, in a properly extended sense of the notion) would allow me to distinguish one of the vases from the other three; no aspect of my present perceptual experience would help, or of my recollections of my wanderings around the cloister.

Now, judgments are constitutively normative acts. Part of Evans' idea, as I understand it, is that whether or not my judgment meets its constitutive norms depends on how things are with the vase I am in fact perceiving, independently of my capacity to descriptively select it from the others. If, in order to be correct, a judgment must just be true, then it is whether or not that specific vase in fact contains my bag that determines whether or not it is correct. If the relevant norms of judgment are

evidence-constrained (if, say, the thinker must *know* the content), then it is whether or not I know, about that vase, that it contains my bag that is relevant. Be this as it may, it is objects themselves, beyond any purely general descriptive means we may resort to in order to have some grip on them, that are relevant to determine whether the constitutive features of *de re* attitudes are met.

Hence, if they do not exist (as they may well not, for all we can tell "from the inside", if the attitudes at stake involve sufficiently difficult epistemic achievements—not just in the case of material objects, but also of some abstract objects), those constitutive features cannot be met. In order to be justified that we are enjoying successful cases of these attitudes, we thus need justification that we are properly related to objects. In the case of material objects, in paradigm cases acquaintance relations (perception, memory, testimony) provide the required justification. We have agreed that those are only paradigm cases, and that discourse can also help entertain successful *de re* attitudes. However, in those cases the justification for entertaining *de re* attitudes towards material objects of the very kind that we have gained through acquaintance in previous cases is also playing an indirect justificatory role.

Cases of failure of reference like (8) suggest a final, additional consideration favouring the 2-D version of Donnellan's take on the contingent *a priori*. Someone who, like Jeshion, wants to defend that it is the very official singular content expressed by (2), (4) and (6) that is both contingent *and a priori* faces a problem. On the 2-D view, what is known *a priori* is not the official object-dependent content; hence, (8) does not pose any problem: there is still a *truth* to be known *a priori*. If it is *a priori knowledge* of that utterance that is claimed, as opposed to merely *a priori* defeasible *justification*, the defender of the contrasting view that the official object-dependent content is known *a priori* will have to envisage true but gappy propositions. This would require a semantics that is technically attainable, but theoretically in need of a justification that is not at all easy to provide.[18] Alternatively, it can be argued that, although acceptance of (8) was *justified a priori*, empirical findings have shown that it is not true. The problem now is that, although there are clear examples of the empirical defeasibility of *a priori* beliefs, it is defeasibility by, say, the testimony of relevant

[18] See Lehmann (2002) for a useful discussion of different kinds of free logics, and the problem they confront to justify the truth-conditions they ascribe to referential sentences. Semantics for free logics should justify the non-validity of rules like, say, existential generalization, and at the same time the truth of sentences like (8), or instances of excluded middle involving non-referring terms. A bivalent proposal like Burge's (1974) achieves this by stipulating that all atomic sentences are false; however, as Lehmann notes (2002, 226), Burge's justification for the stipulation presupposes bivalence, which is at stake once we envisage non-referring terms. Non-bivalent supervaluationist semantics are among the most popular, but they confront a similar problem. Lehmann rightly criticizes a proposal by Bencivenga based on a "counterfactual theory of truth": "Why should truth, which is ordinarily regarded as *correspondence to* fact, be reckoned in terms of what is *contrary to* fact? Why should we reckon that 'Pegasus is Pegasus' *is* true because it *would be* true if, *contrary to fact*, 'Pegasus' did refer?" (2002, 233), concluding, "If supervaluations make sense in free logic, I believe we do not yet know why" (2002, 233). I believe that 2-D accounts, as interpreted here, are in a position to provide the required semantic justification for supervaluationist semantics for free-logics; I hope to elaborate on this elsewhere.

experts that those examples are based on; defeasibility by straightforward empirical findings like those establishing the non-existence of Vulcan is a much more doubtful matter.[19]

These considerations support the view that we should allow for variations in the referent of our complex demonstratives when considering possible worlds as actual, in building up the rows in the relevant propositional concepts. This is all we need to justify considering propositional concepts such as E for the case of instances of the necessary *a posteriori* like (3), thereby having the starting point we need for the Gricean considerations that Stalnaker appeals to so as to understand why the diagonal and not the official content is communicated in those cases; and it also gives us all we need to account for the illusion of possibility along the lines envisaged by Kripke. Nevertheless, the speaker's full justification for understanding the official contents will obviously draw on his linguistic competence. The present proposal is that, in the case of the justification for understanding (3)'s official content, this consists in part in the piece of knowledge that the diagonal proposition for (4) captures. On this view, this is a necessary proposition, on the basis of the meaning-constitution point. We are justified in excluding worlds like *k* in propositional concept *C* as irrelevant, because in those worlds essential semantic properties of the utterances—properties constituting competent understanding—are not kept fixed.[20]

4. The *Locality* and *Context-Dependence* of Apriority

Stalnaker contrasts the metasemantic interpretation of diagonal propositions with another one, which he describes as *semantic*; but he insists that they are complementary in some applications, and my previous cases involving demonstratives might well count among them. The distinction between semantic and metasemantic interpretations of diagonal propositions parallels another distinction he makes, among semantic theories, between *descriptive* and *foundational*: "A descriptive semantic theory is a theory that says what the semantics for the language is without saying what it is about the practice of using that language that explains why that semantics is the right one. A descriptive-semantic theory assigns *semantic values* to the expressions of the language, and explains how the semantic values of the complex expressions are a function of the semantic values of their parts." Foundational theories, in contrast, answer questions "about what the facts are that give expressions their semantic values, or more generally, about what makes it the case that the language spoken by a particular individual or community has a particular descriptive semantics" (1997, 535).

The variations in content represented by the horizontal propositions in a propositional concept depend on a metasemantic interpretation on variations in facts studied by foundational theories, such as for instance causal relations between uses of expressions and things in the world; in a semantic interpretation, they rather correspond

[19] Jeshion (2002b) provides a good discussion.

[20] Discussions with Jim Pryor have helped me to considerably improve a previous version of this section.

to differences determined by facts (other than contents themselves) investigated by descriptive theories, like Kaplan's characters or the kind of reference-fixing descriptive presupposition expressed by (4). In some cases, the semantic interpretation can support applications of the 2-D framework so as to provide the explanatory benefits advertised of it, chief among them that of accounting for the Kripkean phenomena. This notwithstanding, he would presumably point out, in a critical vein, that the "notion of *a priori* that this identification yields is at best a very local and context-dependent one" (this volume, 303, fn. 12).

It is easy to see why he thinks so. Let us start with context-dependence. Consider for illustration a case in which, instead of (3), the speaker uses a simple demonstrative, as in an utterance of (9), relying on what he takes to be the perceptual experiences of his audience to play also the role of the NP that is in (3) a constituent of the complex demonstrative:

(9) If that exists, that is-identical-to-that-evening-heavenly-body

Given that the case is one in which it is taken for granted that the referent of the simple demonstrative is in part fixed by a perceptual experience, presenting it as the brightest morning heavenly body, considerations analogous to those contemplated regarding (4) support ascribing a necessary diagonal proposition to a relevant imaginary utterance of (10):

(10) That is whatever morning heavenly body is demonstrated when that very token of 'that' is uttered, if it exists

Worlds in which the referent of 'that' in (10) is not fixed on the basis of the relevant perceptual experience are of course possible; they are compatible with the knowledge of an otherwise perfectly competent speaker present in the context of the utterance, inattentive to the perceptual circumstances of the case. But those worlds should not count to establish the propositional concept, because not all legitimate presuppositions determining the contribution of the simple demonstrative are in place. It is variations in the referent of the demonstrative when uttered in different circumstances, *keeping fixed what is taken for granted about it* that we are suggesting the diagonal propositions represent. Relative to the context we are considering, then, an account along these lines might count (10) as expressing *a priori* knowledge. There are contexts, however, relative to which it would not express knowledge of that kind, for instance those in which we take into account the presuppositions of the inattentive speaker we just mentioned. This shows that the account, as Stalnaker says, provides a context-dependent notion of *a priori* truth and knowledge.

The case of proper names illustrates the *locality* that Stalnaker ascribes to an account of *a priori* knowledge along the present lines, assuming as he and Kaplan do a Millian view of them. For, in that case, co-referential names like (tokens of) 'Hesperus' and 'Phosphorus' have a constant character, and therefore only the metasemantic interpretation would account for variations in the meaning of the name, so as to allow for worlds like j in propositional concept E above, and thereby contingent diagonal propositions. Hence, assuming the Millian view, the semantic

interpretation can only be invoked locally, in cases like (3)–(10) above; there is no reason to expect a generally valid account of apriority.

However, the Millian view can be contested, and the latter concern at least can thus be discounted. Following Lewis (1983), several philosophers have advanced views according to which the reference of (tokens of) proper names is fixed in part by descriptive metalinguistic information, which speakers know on the basis of their linguistic competence.[21] On some view along these lines, a proper utterance of (11) would express a necessary diagonal proposition, which would constitute knowledge deriving from the semantic competence of speakers confronted, for instance, with related utterances of (1):

(11) Phosphorus is whoever or whatever is saliently called 'Phosphorus' when that token of 'Phosphorus' was uttered, if it exists

The other source of Stalnaker's scepticism about the present account of the treatment of the *a priori* in the 2D-framework would still remain: on this metalinguistic view of proper names, the diagonal proposition expressed by (2) would also count as contingent in some contexts (those in which the relevant reference-fixing information associated with 'Phosphorus' is not common knowledge), and thus it too represents a merely contextual case of *a priori* knowledge similar to the one previously illustrated by means of (10).

I would like to say something to alleviate these doubts. The traditional main concern of epistemologists appears to have been to devise conceptually reductive analyses of the concept of knowledge. Partly due to its lack of success, Williamson (2000) and others have raised serious worries about this enterprise. But even if we still see some point in it, it is clear that there are further tasks for the epistemologist, like making distinctions among kinds of justifications relevant for a clear-headed appraisal of justificatory force. As Wittgenstein's metaphors in *On Certainty* suggest, any sensible distinction between *a priori* and *a posteriori* justifications will be a contextual one, one such that what in a context counts as a proposition justified *a priori*, in another is one justified only *a posteriori* (as (2) and (10) illustrate on the suggested view). But, first, this by itself does not invalidate the significance of the distinction. And, more important, the account highlights propositions whose status as *a priori* knowledge is sufficiently stable across ordinary contexts, as (4) and (11) illustrate among the examples discussed so far; on the present view, the traditional alleged examples of the *a priori* will of course belong in this second group.

In their contribution to this volume (Chapter 3), Byrne and Pryor object to a 2-D proposal like this, along lines to which Millians like Stalnaker would be sympathetic; they make their points concerning Chalmers' views, but I take it that they would think that they also apply to my own. Concerning this latter point about the apriority of (11), they make two objections. The first is that "the metalinguistic proposal imposes unreasonable demands on understanding a word". I already granted

[21] Macià's (2005) proposal includes a nuanced version of this sort of view, which I take to be compatible with the claims I make here.

that there is a burden here for the defender of the account to discharge.[22] However, I cannot see that there is any relevant difference between the burden imposed by the claim of apriority concerning (4), and that concerning (11). Any theoretical elaboration of what it is to understand complex demonstratives will be very far away from what competent speakers by ordinary standards know.[23] The point is that it is at least sufficiently reasonable for the purposes of the present discussion that there should be some aspect of the competence of ordinary speakers (their *personal-level* competence) that is captured by the necessity of the diagonal proposition for (4); the claim is that (11) captures a corresponding aspect, however difficult it is to characterize it in a philosophically satisfactory way.

Byrne and Pryor's second objection is, as I understand it, the one that Frege famously makes in the first paragraph of "On Sense and Reference", in a criticism of metalinguistic accounts of the cognitive significance of identity statements: if we find (1) informative, it is because it gives us astronomical information, not just the information that two names corefer. But I can deal with this on the basis of the previous considerations about the contextuality of *a priori* knowledge. The diagonals of statements like (11) merely capture the most stable aspects of the competence of speakers; there are others, like that captured by the diagonal for (2), and the diagonal for the corresponding statement involving 'Hesperus'. Taking that into account in characterizing the *a posteriori* diagonal for (1), we explain why acceptance of it provides not just uninteresting metalinguistic information, but also astronomical information.

Stalnaker's scepticism about the explanatory potential of the 2-D framework regarding the problems for which it was originally devised is thus unnecessarily defeatist, even granting most of his theoretical assumptions as I think have done. An undeniable difference between the present view and Stalnaker's lies in the rejection of Millianism. However, this does not suffice to equate it with what Stalnaker calls the *generalized Kaplan paradigm*, which "treats a much wider range of expressions as context-dependent: almost all descriptive expressions of the language will have a variable character. While in the original Kaplan theory, it was the content determined that was the thought expressed in the use of an expression, in the generalized theory, it is the character (or the A-intension, or diagonal, that it determines) that is the cognitive value of what is expressed." ("Assertion Revisited", this volume, p. 300.) Although the present proposal agrees with the generalized Kaplan paradigm on the former issue (almost all descriptive expressions of the language will have a variable character), it does not need to agree with it on the latter (it is the character that is the cognitive value); this is where the merely presuppositional role given to the descriptive

[22] See above, fn. 13.

[23] The same applies to simple demonstratives, of course; in fact, I have avoided them for strategic reasons, because they impose more recalcitrant problems. It is theoretically very difficult to reject that speakers have, as part of their competence, the descriptive knowledge of the referent required on the present account for the necessity of the diagonal for (4). It is easier to reject any proposal concerning the corresponding descriptive aspects of understanding simple demonstratives, like the one I would be prepared to make.

reference-fixing information matters. Consequently, the present proposal does not espouse an asymmetry like the one that Stalnaker mentions here:

> One important difference between the two theories is the contrasting roles of the two-dimensional intensions (character, in Kaplan's semantics, propositional concepts in the assertion theory) in the explanation for the fact that an utterance has the content that it has. [. . .] Character precedes content in the order of explanation of the fact that the utterance has the content that it has. But the order is the reverse in the case of the explanation of why an utterance conveys the information that a diagonal proposition represents. [. . .] We explain why the utterance determines the propositional concept that it determines in terms of the content that it has, or would normally have, according to the semantics of the relevant alternative possible worlds. Content (in the various alternative worlds) precedes propositional concept in the order of explanation. The second part of the explanation invokes reinterpretation by diagonalization, but since the diagonal proposition is determined by the propositional concept, the main work of explaining why the utterance conveys the particular content that it conveys is done when we have explained why the utterance determines the propositional concept that it determines. ("Assertion Revisited", this volume, pp. 298–9.)

I agree with Stalnaker (1999, 2) that "it matters what is explained in terms of what"; precisely because of that, I would like to insist that the explanatory priorities he devises in the quoted text are consistent with the present account. What I aim for is an elaboration of the Kripkean suggestions to deal with the epistemological puzzles posed by his views about modality, compatible with my allegiance to them. According to these views, the truth-conditions with respect to possible worlds of the singular claims we make, and their modal status, depend on the objects involved and their objective natures, not on the qualitative ways through which we in the actual world fix reference to them. This is also so on the weaker form of object-dependence for singular contents that I earlier committed myself to.

My disagreement with Stalnaker lies in the fact that he, like other Millians, envisages an asymmetry between indexicals and proper names that I find unwarranted. As far as I can tell, rejecting that asymmetry is compatible with accepting the explanatory priority that Stalnaker wants for singular contents. Notice that, although he develops the argument in the quoted text for an identity statement involving proper names, nothing in the argument itself requires it; the very considerations he appeals to apply also to identities involving indexicals. The argument does not therefore support the Millian asymmetry that distinguishes Stalnaker's view from the one advanced here.

5. Utterance Problems

On the interpretation so far developed, the main aspiration of the 2-D framework is to reconcile the appealing Kripkean metaphysical and semantic views, which envisage substantive *de re* necessities, with the equally intuitive Kripkean views on the epistemology of modality, which in their turn require an explanation for the ensuing illusion that such substantive necessities are contingent. The suggested approach to

attain this goal has been to isolate an *a priori* component in the understanding of the expressions contributing the relevant *res*; diagonal propositions model that component. The concept of *apriority* thus modelled is one along the lines envisaged by Reichenbach, according to which apriority in the relevant cases is a form of analyticity. Given that, as we emphasized at the outset, Kripke's examples are not isolated cases, for the proposal to work it should be established that the distinction between the two sorts of contents can be made in all relevant cases. In particular, one should confront the notorious application by Kripke of his embryonic 2-D suggestions to the mind-body problem, an issue that I am not in a position to discuss here.

Even more ambitious goals for the 2-D framework are embraced by David Chalmers in his contribution to this volume: as he metaphorically puts it, to restore a golden triangle between meaning, reason, and modality allegedly unravelled by Kripke. Less metaphorically, this requires, according to Chalmers, for the two-dimensionalist to sustain a *Core Thesis*, that "for any sentence S, S is *a priori* iff S has a necessary 1-intension". ("The Foundations of Two-Dimensional Semantics", this volume, p. 64.) In contrast with my proposal, here apriority is understood along the lines of earlier philosophical traditions, as an idealized form of knowledge constitutively independent (in some philosophically pertinent sense) of experience. Chalmers classifies different interpretations of 2-D ideas into two main contrasting views, the *contextual* and the *epistemic* understanding of the framework: "the contextual understanding uses the first dimension to capture *context-dependence*. The epistemic understanding uses the first dimension to capture *epistemic dependence*." (*ibid.*) He persuasively argues that the contextual understanding, in any of the different versions he considers, cannot validate the Core Thesis.

How does my proposal fit into Chalmers' taxonomy? If we just attend to his descriptive labels, and the characterization quoted before, it appears not to fully fit, in that *prima facie* it has both contextual and epistemic elements. On the one hand, it makes the semantic features constituting the first dimension dependent on epistemic matters, to wit, those constitutive aspects of understanding I have highlighted before. On the other, in cases like those I have been discussing, it certainly uses the first dimension to capture context-dependence. From the viewpoint developed here, it is only to be expected that the present view does not fully fit into Chalmers' scheme. On the view of context-dependence previously outlined, the reference of context-dependent expressions is constitutively fixed relative to relations involving the relevant tokens; thus, while the identity of the referent itself may well change from accessible epistemic possibility to accessible epistemic possibility, in each of them the referent (if any) stands in the relevant relation to the very same token. If this *token-reflexive* view is correct, thus, any satisfactory epistemic interpretation of the 2-D framework will end up incorporating some features of contextual interpretations.

If this is so, the well-known examples ("utterance problems", concerning examples such as 'language exists', 'someone is uttering', 'I think' and so on) that Chalmers invokes against contextual interpretations—previously mentioned by Evans, Kaplan, and others to similar effects—raise serious questions about the epistemic plausibility

of the token-reflexive view. Here I only have space to encourage the reader to take the 2-D framework seriously when thinking about these examples. Just for illustration, consider the 'language exists' example. Many philosophers would be prepared to take the languages to which generic reference is made here as natural kinds, whose essence might be hidden in the very same sense that the essence of water is. If this is so, there is no problem in ascribing a contingent horizontal content to the claim. The problem, of course, is whether a contingent diagonal, or rather a necessary one, corresponds to the sentence; this is the only issue really at stake.

Concerning this point, I will just make a methodological claim here: this is a delicate theoretical issue, one that we cannot sensibly assume ourselves to be in a position to decide just by appealing to intuition. Philosophical clarification must be provided concerning the knowledge that one must have in order to be able to entertain thoughts about language; and then it must be decided whether or not it is to be expected that any (relevant) thinker does have that knowledge. It is not implausible that an account of the *a priori* along Reichenbachian lines ends up deciding these theoretical issues in such a way that the claims at stake should have necessary diagonals. A view like this cannot be dismissed merely by claiming that the intuitions that one has (on this highly theoretical issue) go against the view.

In his paper, Chalmers says that on the epistemic understanding of the framework (albeit not on the contextual understanding) "there is an intuition . . . that 'I am not uttering now' is not false *a priori*, so that there are epistemic possibilities in which it is true" ("The Foundations of Two-Dimensional Semantics", this volume, p. 119). This is precisely the kind of claim that I want to resist. The concepts of *apriority* and *epistemic possibility* are philosophical, highly theoretical ones. I do not dispute that, after long exposure to philosophical discussions, one can develop the sort of intuitions whose existence Chalmers asserts. The question is what methodological relevance appeal to them has in philosophical discussions such as this. I would say, the same as that of intuitions of highly skilled linguists about the grammaticality of very complex sentences, on which the correctness of grammatical theories crucially turn, about which, when questioned, ordinary speakers simply stare blankly: namely, none. Whether or not a philosophically useful concept of apriority will make claims like 'there is thinking going on' (with diagonals essentially equivalent to that for the claim whose denial Chalmers contemplates) *a priori* is up for grabs: it is not the sort of issue to be decided by an appeal to intuition.

References

Bealer, George (1999). "A Theory of the *A priori*", *Philosophical Perspectives* 13, 29–55.

Bonjour, Laurence (1998). *In Defense of Pure Reason*, Cambridge: Cambridge University Press.

Burge, Tyler (1974). "Truth and Singular Terms", *Noûs* 8, 309–25.

——(1993). "Content Preservation", *Philosophical Review* 102, 457–88.

Burgess, John P. (1998). "Quinus ab Omni Nævo Vindicatus", *Canadian Journal of Philosophy: Meaning and Reference*, sup. vol. 23, A. Kazmi (ed.), 25–65.

Chalmers, David (forthcoming). "Two-Dimensional Semantics", in E. Lepore and B. Smith (eds.), *Oxford Handbook of Philosophy of Language*, Oxford: Oxford University Press; 13/1/2005, http://consc.net/papers/twodim.html.

Dever, Josh (2001). "Complex Demonstratives", *Linguistics and Philosophy* 24, 271–330.

Donnellan, Keith (1979). "The Contingent *A priori* and Rigid Designation", in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, Minneapolis: University of Minnesota Press, 45–60.

Evans, Gareth (1982). *The Varieties of Reference*, Oxford: Clarendon Press.

Fine, Kit (1995). "Ontological Dependence", *Proceedings of the Aristotelian Society* 95, 269–90.

García-Carpintero, Manuel (1998). "Indexicals as Token-Reflexives", *Mind* 107 (1998), 529–63.

____ (2000). "A Presuppositional Account of Reference-Fixing", *Journal of Philosophy* 97 (3), 109–47.

Glanzberg, Michael and Siegel, Susanna (forthcoming). "Presupposition and Policing in Complex Demonstratives", *Noûs*.

Jeshion, Robin (2001). "Donnellan on Neptune", *Philosophy and Phenomenological Research* 63, 111–35.

____ (2002a). "Acquaintanceless *De Re* Belief", in Joseph Keim Campbell, David Shier, and Michael O'Rourke (eds.), *Meaning and Truth: Investigations in Philosophical Semantics*, New York: Seven Bridges, 53–78.

____ (2002b). "The Epistemological Argument Against Descriptivism", *Philosophy and Phenomenological Research* 64, 325–45.

Kaplan, David (1989). "Demonstratives", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 481–563.

King, Jeffrey (2001). *Complex Demonstratives: A Quantificational Account*, Cambridge, Mass.: MIT Press.

Kripke, Saul (1980). *Naming and Necessity*, Cambridge, Mass.: Harvard University Press.

Lehmann, Scott (2002). "More free logic", in Gabbay, ed. *Handbook of Philosophical Logic*, 2nd edn, Vol. 5, Kluwer-Academic Publishers, 197–259.

Lewis, David (1983). "Individuation by Acquaintance and by Stipulation", *Philosophical Review* 92, 3–32.

Macià, Josep (2005). "Proper Names: Ideas and Chains", in M. Ezcurdia, C. Viger, and R. Stainton (eds.), *New Essays in Language and Mind. Canadian Journal of Philosophy*, Supplementary Volume.

McLaughlin, B. P. and Tye, M. (1998). "Is Content Externalism Compatible with Privileged Access?", *Philosophical Review* 107, 349–80.

Peacocke, Christopher (1993). "How Are *A priori* Truths Possible?", *European Journal of Philosophy* 1, 175–99.

Pérez-Otero, Manuel and García-Carpintero, Manuel (1998). "The Ontological Commitments of Logical Theories", *European Review of Philosophy* 4, 157–82.

Perry, John (1997). "Indexicals and Demonstratives", in D. Wright and B. Hale (eds.), *A Companion to the Philosophy of Language*, Oxford: Blackwell, 586–612.

Reichenbach, Hans (1920). *Relativitätstheorie und Erkenntnis Apriori*, Berlin: Springer. English translation as *The Theory of Relativity and A priori Knowledge*, Berkeley and Los Angeles, University of California Press (1965), from which I quote.

Soames, Scott (2003). *Philosophical Analysis in the XXth Century, Vol. 2: The Age of Meaning*, Princeton, N.J.: Princeton University Press.

——(2005). *Reference and Descriptions: The Case against Two-Dimensionalism*, Princeton, N.J.: Princeton University Press.

Stalnaker, Robert (1978). "Assertion", in P. Cole (ed.) *Syntax and Semantics* 9, New York: Academic Press, 315–32; also included in Stalnaker (1999), from which I quote.

——(1997). "Reference and Necessity", in C. Wright and B. Hale (eds.), *A Companion to the Philosophy of Language*, Oxford: Blackwell, 534–54.

——(1999). *Context and Content*, Oxford: Oxford University Press.

Williamson, Timothy (2000). *Knowledge and Its Limits*, New York: Oxford University Press.

8

Phenomenal Belief, Phenomenal Concepts, and Phenomenal Properties in a Two-Dimensional Framework

Martine Nida-Rümelin

1. Phenomenal Belief, Phenomenal Concepts, and Phenomenal Properties

Peter, who is looking at the cloudless sky during the day, and Eve, who is looking at a painting of Yves Klein, have something in common. They both have a visual experience that has a common feature with respect to the color sensation. They are both having a blue sensation. The property of having a blue sensation is a paradigmatic example of phenomenal properties. Phenomenal properties are often conceived of as properties of inner events or processes. I prefer to think of phenomenal properties as properties of sentient beings.

A person who never had color experiences may have a concept of the property of having blue experiences (acquired by talking with sighted people or by reading books) but she does not have a *phenomenal* concept of having blue experiences. Phenomenal concepts are acquired on the basis of one's own experiences of the relevant kind. Phenomenal concepts of having color experiences of particular kinds are acquired on the basis of one's own color experiences. The question of how to account for the relation between phenomenal concepts and phenomenal properties is at the center of the current debate about the ontological status of consciousness. Now every account of phenomenal concepts within some proposed theoretical framework must be tested against our intuitive pre-theoretical understanding of the notion of phenomenal concepts. It is therefore important to sharpen our intuitive understanding of what it is to have a phenomenal concept of a particular phenomenal property before entering the theoretical debate about the appropriate theoretical account of phenomenal concepts and their relation to phenomenal properties.

Phenomenal concepts are involved in what I call phenomenal belief. The best and maybe even the only way to get a clear intuitive understanding of what it is to have a phenomenal *concept* of a phenomenal property is to get a clear intuitive understanding of what it is to have a phenomenal *belief*—a belief involving a phenomenal concept.

2. Phenomenal Belief

When Frank Jackson's famous Mary leaves her black and white room (she has been spending her entire life in that room and never had any kind of color experience) and when she finally looks at the blue sky, she learns something new about the color experiences of other people.[1] She acquires the phenomenal belief that the sky appears in that particular color (blue) to normally sighted people. When she finally sees the sky, Mary takes two steps at once: *she acquires the phenomenal concept of having blue experiences and she forms a correct belief involving this new concept*. To see that there are two epistemic steps involved (acquisition of a phenomenal concept, and the formation of a belief involving the concept) it is useful to consider Marianna's case:[2] Marianna, like Mary, has always been living in a black-and-white environment. One day, however, the house she has been living in so far is radically changed. She finally gets acquainted with colors: the tables, chairs, etc. are painted in many different colors, the walls are now decorated with abstract paintings. But she does not see any bananas, tomatoes or pictures of landscapes. She does not see any of those objects the color of which she already knows under some of her previously acquired concepts. While looking at four different slides one after the other (a blue one, a green one, a yellow one and a red one) and while enjoying the color experience they provoke, Marianna wonders which of the slides causes in herself the phenomenal kind of color sensation caused in normally sighted people when looking at the cloudless sky. She thereby considers a question that she could not have considered before. She is now able to make new epistemic mistakes that she could not have made before. After having thought about the question for a while, Marianna may well form the new belief that the sky appears to normally sighted people like the red slide appears to her. She then entertains a false belief about how the sky appears to normally sighted people. Her false belief involves the *phenomenal* concept of having red sensations. On the basis of her acquaintance with colors Marianna has acquired the epistemic capacity to ask new questions and to make new mistakes. This is explained by the fact that she has new concepts: phenomenal concepts of phenomenal properties. The phenomenal concept of having red sensations is the concept used by Marianna in this particular false belief to attribute a particular kind of sensations to normal people when looking at the cloudless sky.

The acquisition of phenomenal concepts involves having relevant phenomenal properties oneself. But having or having had a particular phenomenal property is neither necessary nor sufficient for the acquisition of the phenomenal concept of that particular property. It is not sufficient because a sentient being may experience a particular color without forming the phenomenal concept of the property of having that kind of color experience. A sentient being has the phenomenal concept of the property of having a particular kind of color experience only if it is able to

[1] See Jackson (1982).

[2] This case is discussed at length and used to introduce the distinction between phenomenal and non-phenomenal beliefs about experiences in my papers (1996) and (1998).

attribute that property (under that concept) to another sentient being (only if it is able to consider the question whether and form the belief that another being has that particular kind of experience). It is possible (for example, for an animal) to have a particular color experience without being able to attribute having this kind of experience to another being. Therefore, acquaintance with a particular kind of experience does not entail forming a concept of the property of having that kind of experience. The other direction does not hold either (or at least it is not obvious that it does). A person who never had an experience of orange might be able to form the concept of having an experience of orange on the basis of her acquaintance with red and yellow.

3. The Two-Dimensional Framework

According to the view proposed in this paper, what makes phenomenal concepts different from most other concepts is the intimate relation between the concept and the property expressed by the concept. This intimacy can be formulated like this: *in the particular case of phenomenal concepts to understand the concept involves grasping the corresponding property.* But what is it to grasp a property and what is it to understand a concept? And how if at all is it possible to argue for the claim just formulated? I will not be able to give a satisfying and complete answer to these questions. But it does seem clear to me that a first sketch of how these questions have to be addressed can be formulated quite naturally within the two-dimensional framework in one of its possible and well-known interpretations advocated by David Chalmers.[3]

Since this is a book on two-dimensionalism, I will only very briefly recall the intuitive interpretation of the two-dimensional function and formulate a few standard definitions. I will use the framework to describe concepts and not to describe the meaning of linguistic expressions.[4] The two-dimensional function F_C describes the concept C where

$$F_C(<w1, w2>) = E$$

[3] This interpretation is described and defended in Chalmers (1996). I will not make use in the present paper of his later elaborations.

[4] In the discussion of phenomenal concepts this detail is important. One may plausibly argue that the secondary intension of the linguistic expression "person P has a green sensation" depends on the kind of sensation caused in normally sighted people by paradigmatically green objects. In different possible worlds paradigmatically green objects cause different kinds of color sensations in people with normal vision relative to the standards of that world. Therefore, the different possible secondary intensions of "P has a green sensation" do not coincide. However, as will be argued below, the two-dimensional function that describes the phenomenal concept of having green sensations does have identical possible secondary intensions. Therefore, in this particular case, the concept and the corresponding linguistic expression are appropriately described by *different* two-dimensional functions. Presupposing the plausible assumption that a concept C is captured in a linguistic expression E only if the two-dimensional function describing the concept C and the two-dimensional function describing the meaning of the expression E are identical, we get the result that phenomenal concepts are not captured in public language. For an elaboration of this point see my paper (2003a).

has the following intuitive interpretation: The extension of the concept C in the counterfactual world w2 would be E if w1 were the actual world. Just one example: let w1 be a world where Joschka Fischer is the President of France in 2004. Then the extension (reference) of the concept associated to the rigid definite description "the president of France in 2004" in every w2 would be Joschka Fischer if w1 were the actual world.

Definition 1: The primary intension PI_C of a given concept C is defined as follows:

For every w: $PI_C(w) = F_C(<w,w>)$

So the primary intension of a concept C tells us for every possible world w what would be the extension of the concept C in the world w if w itself would be the real world. In other words, the primary intension describes how the extension of the concept in the real world depends on features of the real world.

Definition 2: The secondary intension SI_C of a given concept C is defined as follows:

For every w: $SI_C(w) = F_C(<w_{actual}, w>)$

where w_{actual} is the actual (the real) world.

So the secondary intension of a concept C delivers the extension of the concept C in every counterfactual world w given that the real world has the relevant features it really has. In the case of a property concept the secondary intension describes what features are necessary and sufficient for an entity to fall under the concept in counterfactual circumstances given that those entities falling under the concept in the real world have the features they actually have. In many cases the secondary intension depends on features of those entities falling under the concept in the real world that we still do not know. In general, therefore, conceptual knowledge is not sufficient to know the secondary intension of a given concept (but there are exceptions).

For what follows it will be useful to introduce the notion of a possible secondary intension relative to a world w.

Definition 3: The possible secondary intension of a concept C relative to the possible world w1 $SI_{C,w1}$ is defined as follows:

For all w2: $SI_{C,w1}(w2) = F_C(<w1, w2>)$

The secondary possible intension of a given concept C relative to a world w is what would be the secondary intension of C if w were the actual world.[5]

4. Grasping Properties

To grasp a property (or to grasp the nature of a property) is to know what is essential or constitutive of having that property. To know what is essential of having a particular property expressed by a given property concept C is to know the conditions that

[5] In the present paper indexicals and demonstratives will not play any role. Therefore I skipped centered worlds in the introduction of the two-dimensional function.

are necessary and sufficient for something to fall into the extension of C in all possible factual or counterfactual circumstances. Formulated within the two-dimensional framework we may say in a first approximation: to grasp a property is to have implicit knowledge of its secondary intension (or, for short: of its *counterfactual extension*). This view about what it is to grasp a property may be motivated by the observation that controversies about the nature of the property expressed by a given concept (for example by the concept of having a blue sensation) are typically controversies about what condition is necessary and sufficient for an individual in counterfactual circumstances to fall under the concept. The functionalist believes that to occupy a specific causal role is essential or constitutive of having the property expressed by the phenomenal concept of having a blue sensation. He or she thereby claims that to fall under the phenomenal concept of having a blue sensation it is necessary and sufficient for a being in counterfactual circumstances to occupy causal role R. Therefore controversies about the functionalist claim typically take the form of controversies about whether or not a being in counterfactual circumstances may be in the extension of the phenomenal concept at issue and yet not occupy causal role R and vice versa (this is of course why inverted spectrum, absent qualia and inverted earth scenarios are relevant to the debate about what is constitutive or essential for having the property at issue). Another intuitive way to see why grasping the nature of a property expressed by a concept C is to have implicit knowledge of its secondary extension is this: those entities falling under a concept C in the real world may share features that are not essential for having the property expressed by C. What features a being may lose and still fall under C is an information contained in the secondary intension of C. To have this information is to know what is essential for the property expressed by C.

These reflections make it plausible to identify grasping a property P with implicit knowledge of the secondary intension of some concept that expresses P. But the proposal is not without problems. One may wish to say that the geometrical property of being a triangle with three equal angles and of being a triangle with three equal sides is not the same property although the two concepts C1 (the concept of the property of having three equal angles) and C2 (the concept of the property of having three equal sides) that express these properties have the same secondary intension. What should we say then about a person who knows that for a triangle to fall under C1 it is necessary and sufficient to have three equal sides but does not know that to fall under C1 it is necessary and sufficient to have three equal angles. If we accept that the properties expressed by C1 and C2 are different, then we might wish to describe this as a case where the person at issue has not yet grasped the property expressed by C1. But she has implicit knowledge of the secondary intension of C1. So we have a counterexample to the explication of grasping properties expressed by concepts.

One way to react to this problem is to deny that C1 and C2 express different properties. I'm not sure that this reaction is appropriate. Another possibility is to accept the counterexample and to weaken the proposal accordingly. According to the new proposal we should say something like this: If a person has grasped the property expressed by a concept C then she has implicit knowledge of C's secondary intension and if she has implicit knowledge of C's secondary intension then she

has at least quasi-grasped the property expressed by C in the following sense: there is some other concept C′ that has the same secondary intension as C and she has grasped the property expressed by C′. The difference between grasping and quasi-grasping properties expressed by concepts will, however, not be of any importance in the present context. As far as I can see the specific problem of hyperintensionality is irrelevant for our understanding of phenomenal concepts, phenomenal properties and the relation between them.

A few points of clarification: (a) the assertion "the person P grasps the property expressed by the concept C" may change its truth value when "concept C" is replaced by a term that refers to a concept expressing the same property. The assertion in the sense here intended implies that the property at issue is grasped via the concept C. (b) I presuppose that the grasping of properties always is a grasping via *some* property concept. We can, however, introduce a notion of "x grasps the property P" (a formulation that does not mention a particular concept of P) by quantifying over concepts: x grasps property P iff there is some concept C such that x grasps the property expressed by C. (c) I presuppose an intuitive notion of what it is for a concept to express a particular property. The intuitive notion may be clarified within the two-dimensional framework. Let us assume that properties can be represented by functions from possible worlds into extensions. Then the following definition seems to me to capture the intuitive notion at issue: A given concept C expresses the property P iff the function that represents the property P coincides with the secondary intension of the concept C.

5. Understanding Concepts

According to the interpretation of the two-dimensional framework here presupposed the primary intension of a concept is knowable *a priori* while in general the determination of its secondary intension requires factual knowledge. This may invite the conclusion that to understand a concept is to have implicit knowledge of its primary intension. But this would be a mistake. It is part of, for example, our concept of water that nothing in counterfactual circumstances falls into the extension of the concept of being water unless it shares its chemical composition (or, more general: its hidden scientific nature) with the liquid falling under the concept in the real world. The way the counterfactual extension (secondary intension) depends on features of entities falling under the concept in the real world is represented in the two-dimensional function as a whole. Therefore understanding a concept requires more than implicit knowledge of its primary intension. *Understanding a concept may be described as implicit knowledge of the corresponding two-dimensional function as a whole.*

According to the model here proposed grasping a property expressed by a concept C is implicit knowledge of its secondary intension and understanding a concept C is implicit knowledge of the corresponding two-dimensional function F_C. Using this model we can say what is required *in addition* to the understanding of a property concept C for the grasping of the corresponding property. A person who understands the concept C needs to know enough about the real world such that the secondary intension of the concept does not depend any more on any still unknown feature of

the real world. Let W be the set of possible worlds that represents what P believes about the real world (this is to say: according to what p believes, every world in W and only worlds in W can be the real world). In order for p to grasp the property expressed by C, W must be such that for all worlds w and w′ in W the possible secondary intensions SI_w and $SI_{w'}$ coincide. This is meant to capture the intuitive idea that *there is no possible discovery for p to make about the real world such that the counterfactual extension of her concept depends on the result of that discovery*.

6. Actuality-Independent Concepts and Grasping Properties

The above reasoning motivates the claim that a person grasps the property expressed by a property concept (understands its nature) if she has an actuality-independent concept of that property, where actuality-independent concepts are defined as follows.

Definition 4: A concept C is *actuality-independent* iff the corresponding two-dimensional function F_C fulfills the following condition:

$$\forall w \, \forall w' \quad SI_w = SI_{w'}$$

Concepts that are not actuality-independent will be called actuality-dependent. In the case of an actuality-independent concept having (understanding) the concept is sufficient for grasping the property expressed by the concept. Who understands the concept (has implicit knowledge of its corresponding two-dimensional function) thereby knows its counterfactual extension (has implicit knowledge of its secondary intension). However, a person need not have an actuality-independent concept of a property in order to grasp the property. She may have an actuality-dependent concept of the property at issue but given her background knowledge the counterfactual extension of her concept may not depend any more on any still unknown feature of those entities that fall under her concept in the real world. Presupposing that the concept of being composed of H_2O is actuality-independent (although this may be doubted) a person who understands the concept of water and knows that water is composed of H_2O is an example. Although the concept of being water is not actuality-independent, we should describe such a person as a person who has grasped the property expressed by the concept of being water given her background knowledge. This and similar examples can motivate the following definitions.

Definition 5: The concept C is *actuality-independent relative to the set of possible worlds W* iff

$$\forall w \, \forall w' \, (w \in W \, \& \, w' \in W \rightarrow SI_w = SI_{w'})$$

Definition 6: Let W_p represent the background knowledge of a given person P. Then the person P grasps the property expressed by the concept C iff C is actuality-independent relative to W_p.

The set of possible worlds that represents her background knowledge contains only those worlds and all those worlds that could be the real world given what P *correctly*

believes. It is necessary to add "correctly" (and to relativize in the above definition to background *knowledge* and not just to background *belief*) since the definition is intended to capture *real* grasping (grasping the real nature of a property) and not just *apparent* grasping.

A person who grasps a property need not have an actuality-independent concept of that property, but a person who has an actuality-independent concept of a property grasps the property expressed by that concept for she then trivially satisfies definition 6.

7. The Special Status of Phenomenal Concepts

A person who understands a phenomenal concept thereby grasps the property it expresses. This natural intuitive idea can now be explained and justified within the proposed framework. It suffices to show that phenomenal concepts are actuality-independent. This claim is found in David Chalmers' well-known thesis that in the case of phenomenal concepts the primary intension coincides with the secondary intension (more precisely: with any of its possible secondary intensions). It is easy to show that a concept is actuality-independent iff it has this property.[6] To justify the thesis that phenomenal concepts are actuality-independent we need to give intuitive support to the idea that in the special case of phenomenal concepts the counterfactual extension of the concept does not depend on any feature of the real world.

At first sight the claim may appear to be obviously wrong. Of course there is a dependence of the counterfactual extension of the concept of having a blue sensation on features of the real world: sentient beings in counterfactual circumstances fall under the concept of having a blue sensation just in case they have a sensation that has (with respect to color) *the same subjective character as those people falling under the concept in the real world*. If people falling under the concept of having blue sensations in the real world (sentient beings in the actual extension of the concept) had sensations of yellow, then a sentient being in counterfactual circumstances would fall into the extension of the concept iff it had a yellow sensation. So what falls under a phenomenal concept in counterfactual circumstances depends on the qualitative character experienced by those who fall under the concept in the real world. In general, the counterfactual extension of a concept depends in ways determined by the concept on features of those entities falling under the concept in the real world. In the case of the concept of being water the counterfactual extension depends on the chemical structure of the liquids falling under the concept of being water in the real world. *In*

[6] See Chalmers (2002). Proof of the claim that a concept is actuality-independent iff its primary intension is identical with any of its secondary intensions: C is actuality-independent iff

$\forall w1 \forall w2$: $SI_{C,w1} = SI_{C,w2}$ iff
$\forall w1 \forall w2\ \forall w\ F_C(<w1,w>) = F_C(<w2,w>)$ iff
$\forall w1 \forall w2\ \forall w\ F_C(<w1,w>) = F_C(<w2,w>) = F_C(<w,w>)$ iff
$\forall w1 \forall w2$: $SI_{C,w1} = SI_{C,w2} = PI_C$.

the case of phenomenal concepts the counterfactual extension depends on the phenomenal character experienced by those sentient beings that fall under the concept in the real world. So there *is* a dependence of counterfactual extension on features of what falls into the real extension in the case of phenomenal concepts as well. Therefore, the counterfactual extension (the secondary intension of phenomenal concepts) should depend, it may seem, on which world is taken to be the actual world (in other words: there are, it may seem, different possible secondary intensions). If w is a world where those people falling under the concept of having a blue sensation have blue sensations and w′ is a world where those people falling under the concept of having a blue sensation have yellow sensations, then the corresponding secondary intensions SI_w and $SI_{w'}$ are different. Therefore—one may be tempted to conclude—the phenomenal concept of blue sensations is not actuality-independent.

But the conclusion is wrong. *There are no worlds where the actual extension of the concept of having a blue sensation contains all and only people having yellow sensations. Phenomenal concepts are individuated by the type of phenomenal experience they pick out in the real world*. A phenomenal concept that has in its extension all and only people with yellow sensations is *ipso facto* the phenomenal concept of *yellow* sensations. This leads us to an important disanalogy between the concept of being water and phenomenal concepts. There are possible worlds w such that our concept of being water would pick out all and only liquids composed of XYZ if w were the actual world. The chemical structure is, let us say, the essential feature of being water. Then we may say this: there are possible worlds w such that the real extension of our concept of being water would contain liquids with a different essential feature if w were the real world. But there are no possible worlds w such that the extension of our concept of having a blue experience would contain people having experiences with another phenomenal character. Having an experience of a particular phenomenal character is the essential feature of having blue experiences. So we may contrast the case of phenomenal concepts with the case of the concept of being water in this way: *In the case of the concept of being water there are possible worlds w such that the elements of the actual extension of the concept would be different with respect to the essential feature (chemical structure) associated with the concept if w were the actual world. In the case of phenomenal concepts there is no world w such that the elements of the extension of the phenomenal concept would be different with respect to the essential feature associated to the concept if w were the actual world.* Therefore, in the case of being water the possible secondary intensions associated to different possible worlds are different while in the case of phenomenal concepts they are not: *in the case of phenomenal concepts the elements in the actual extension of the concept necessarily share the essential feature associated with the concept.* This is why all possible secondary intensions coincide. A concept with identical possible secondary intensions is an actuality-independent concept. So phenomenal concepts are actuality-independent.

We now can formulate and justify the idea that in the particular case of phenomenal concepts understanding the concept implies grasping the property it expresses: A person who has an actuality-independent concept C of a property grasps the property expressed by the concept C. Therefore, a person who has the actuality-independent

phenomenal concept of having blue experiences thereby grasps the property of having blue experiences (knows what is essential for having the property it expresses).[7]

8. A Transparency Principle

Is it possible to grasp one and the same property via two different concepts? We should, I think, allow for this possibility. You can grasp the property of having sensations of orange via the phenomenal concept of having sensations of orange but also via the concept of having a sensation of a color phenomenally composed of red and yellow. If you have enough chemical background knowledge you may grasp the property of being water via your concept of being water but you can also grasp the same property via your concept of being composed of H_2O. But in these cases it will be possible for you to see that you have grasped one and the same property in two ways. If the idea of grasping properties makes any sense at all, then it should in principle be possible for a person who has grasped one and the same property in different conceptual ways to find out that she thereby has grasped the same property. Identical properties are necessarily coextensional (they are represented by the same secondary intension). So we may formulate the intuitive idea just mentioned like this: *A person who has grasped a property P by two different concepts C1 and C2 is in principle capable to find out that C1 and C2 are necessarily coextensional.* This *transparency principle* may be seen as a partial explanation of what it is to grasp a property via a concept.[8]

It is interesting but also a little troubling to see that the transparency principle is trivial within the two-dimensional framework given the definitions I have proposed and presupposing the following further translation of an intuitive locution into "the two-dimensional framework": A person can in principle know that C1 and C2 are necessarily coextensional just in case she has background knowledge H such that C1 and C2 have identical possible secondary intensions for every world w compatible with her background knowledge H (that is $SI_{C1,w} = SI_{C2,w}$ for every $w \varepsilon H$, where H represents her background knowledge). Now to say that the person has grasped the property expressed by C1 and that she has grasped the property expressed by C2 is to say that she has background knowledge H such that C1 and C2 are both actuality-independent relative to H. To say that they are both actuality-independent relative to H implies that for both concepts the possible secondary intensions still compatible with H coincide with the real secondary intension (for all $w \varepsilon H$ $SI_{C1,w} = SI_{C1}$ and $SI_{C2,w} = SI_{C2}$). But the secondary intensions of these concepts are identical since the two concepts express the same property. Therefore, a person who has grasped both concepts thereby has background knowledge H such that C1 and C2 have identical

[7] For a critical elaboration of this reasoning see my (2006), sections 10 and 11.

[8] The choice of the term 'transparency principle' is inspired by Stephen White (2005), who uses the term 'the intuition of transparency' in connection with the thesis that Loar (1997) describes as: "... if two concepts conceive of a property essentially, neither mediated by contingent modes of presentation, one ought to be able to see a priori—at least after optimal reflection—that they pick out the same property." (p. 600.)

possible secondary intensions for every world w compatible with H. So a person who grasps both concepts knows that they are necessarily coextensional.

Given the translations proposed—and these translations are quite natural once the intuitive interpretation of the two-dimensional framework here presupposed is accepted—the transparency principle is a logical triviality within that framework. This may be taken to be good news for the philosopher who wishes to justify the transparency principle. But it may also be taken to be bad news for the framework in its presupposed interpretation. The transparency principle appears to be a substantial assumption. Therefore, or so one may argue, it should not turn out a logical triviality within a framework used to describe the relation between concepts and properties. If it does so turn out, this shows that there are substantial and potentially controversial assumptions built into the conceptual framework at issue. I am not sure what to conclude from these observations. In the following I will presuppose the truth of the transparency principle.

9. Mary's Epistemic Progress Revisited

A natural reaction to Mary's story may be formulated like this: Mary before her release does in some sense know of her normally sighted friend Peter that he has blue sensations when looking at the cloudless sky during daytime, but she has no full understanding of the property she ascribes to Peter in that belief. She has not yet grasped the property of having blue sensations. She only has a deferential concept of having blue sensations but she has not grasped what having blue sensations consists in.

Now it follows from what has been said so far that Mary does grasp the property of having blue sensations once she has acquired the phenomenal concept of having blue sensations. But is does not yet follow from what has been said that she does *not* grasp that property before her release by some *other* concept. The identity theorist may object that Mary may well have a physical-functional concept of having blue sensations such that she can *grasp* what having blue sensations consists in via that physical-functional concept.

Note that the identity theorist can but need not reply in that way. His claim is that having blue sensations *is* a physical-functional property. He need not claim that it is possible to *grasp* the property of having blue sensations via any physical-functional property. There are still two possible views compatible with his identity claim that do not imply the stronger thesis that the property of having blue sensations can be grasped via physical-functional concepts: (a) the identity theorist may insist that—although the property of having blue sensations cannot be grasped via some physical-functional concept—there is some physical-functional concept that *expresses* the property of having blue sensations (there is a physical-functional concept C such that the secondary intension of that concept coincides with the secondary intension of the phenomenal concept of having blue sensations) or (b) he may claim that although the property of having blue sensations is a physical-functional property there is no physical-functional *concept* that expresses that property.[9] Identity theorists

[9] Galen Strawson (1999) may be interpreted as accepting (a). Flanagan (1992) can be interpreted either as accepting (a) or as accepting (b).

who take one of these lines have to deny what appears quite plausible at least at first sight: Every physical property can in principle be *grasped* by some physical-functional concept. The two views just mentioned try to integrate the following three claims: (a) what having blue sensation consists in can be grasped via a phenomenal concept and (b) what having a blue sensation consists in cannot be grasped via a physical-functional concept, but (c) having a blue sensation *is* to be in a particular physical-functional state. One may suspect that an identity theorist of this type has conceded too much to the dualist and that his position may well turn out to be inconsistent on further reflection and on the basis of further premises that are hard to deny.[10]

There is then some motivation for the identity theorist to insist that the property of having blue sensations *can* be *grasped* via a physical-functional concept that is available to Mary before her release. Reformulated in the framework here proposed: There is a physical-functional concept C such that relative to Mary's complete physical knowledge H about human color vision, C is an actuality-independent concept relative to H and the secondary intension of C coincides with the secondary intension of the phenomenal concept of having blue sensations. The dispute between the dualist and the identity theorist thus should turn to this particular question. Is it at all plausible that there is a physical-functional concept that fulfills these two conditions? How can we decide the issue?

At this point the transparency principle may be of some help. Let us suppose that Mary acquires the phenomenal concept of having blue sensations in the particular way Marianna does (she does not see the sky or any other paradigmatically blue objects). There has been some discussion about whether Mary will be able to find out which of her phenomenal concepts expresses the property called "having blue sensations". In other words: If C is Mary's physical-functional concept of having blue sensations (let us suppose for the moment—against the dualist thesis—that there is such a concept), is it possible for Mary to find out that C and her new phenomenal concept of having blue sensations have the same extension in the real world? Some have argued that Mary would be able to find that out given her rich neurophysiological background.[11] But if we accept the transparency principle and if we wish to judge whether Mary can *grasp* the property of having blue sensations via some physical-functional concept, then we have to consider a related but different question: will it be possible at this point for Mary to find out that the physical-functional concept C and the phenomenal concept of having blue sensations are *necessarily* coextensional? Necessity in this context is to be interpreted as metaphysical necessity. (Nomological necessity is not sufficient since different properties may be coextensional in every nomologically possible world.)[12]

Some identity theorists may be tempted to argue in a quite simple way for the claim that Mary *will* be able to find out that her concept C and the phenomenal concept

[10] For a discussion of this view see my papers (2004) and (2006).

[11] See Dennett (1984) and Hardin (1992).

[12] Hardin's arguments in his papers (1987) and (1992) may be interpreted as support for the claim that Mary will be able to find out that C and the phenomenal concept of having blue sensations are necessarily coextensional in the sense of metaphysical necessity. For a critical discussion of his argument see my paper (1999).

of having blue sensations are necessarily coextensional. They will rely on the analogy with the water/H_2O case and will say something like this: we have discovered the identity of being water with being H_2O. Once we have discovered the identity of these properties, we thereby have discovered that the concept of being water and the concept of being H_2O are necessarily coextensional since there simply is no metaphysically possible world where something has property P and does not have property Q if P and Q are one and the same property. The same kind of reasoning, they may continue, is available to Mary. After her release, she will find out that having the physical-functional property expressed by her concept C and having the property expressed by her new phenomenal concept of having blue sensations is one and the same property. She thereby has discovered that the concept C and the phenomenal concept of having blue sensations are necessarily coextensional.

But this simple argument, or so I claim, is not convincing. The claim that the properties expressed by a concept C1 and a concept C2 are identical is only justified in a case where the following modal claim is justified: C1 and C2 have the same extension in every metaphysically possible world. But the latter modal claim is never the result, or so one may argue, of empirical knowledge *alone*. Rather, it involves empirical knowledge *and* conceptual knowledge. *It involves, in particular, conceptual knowledge about what kind of features are candidates for what is essential for falling under the concept at issue.* In the case of being water and being H_2O we can conclude on the basis of empirical results that the two concepts are necessarily coextensional. But we can do so only on the basis of the additional *conceptual* claim: to fall under the concept of water a liquid in counterfactual circumstances must share its chemical composition with what falls under the water-concept in the real world. We obtain the result that the concept of being water and the concept of being H_2O are necessarily coextensional by the following argument.

Premiss P1 (empirical result): The concept of being water and the concept of being composed of H_2O have the same extension in the real world.

Premiss P2 (conceptual insight): A liquid in counterfactual circumstances falls under the concept of being water iff it has the same chemical composition (hidden scientific nature) as the liquids falling under the water-concept in the real world.

Therefore:

(1) A liquid in counterfactual circumstances falls under the water-concept iff it is composed of H_2O. (From P1 and P2.)

Therefore:

(2) The concept of being water and the concept of being composed of H_2O are necessarily coextensional. (From (2), presupposing that something in counterfactual circumstances falls under the concept of being composed of H_2O just in case it is composed of H_2O.)

Given this result we are allowed to identify the two properties. This identity claim is not an empirical result. It is obtained on the basis of an *empirical* premise (P1) and a *conceptual* premise (P2).

Is there a parallel argument available to Mary that justifies the claim that the property expressed by C and the property expressed by the phenomenal concept of blueness is one and the same? An analogous argument would have to take the following form.

Premiss P1′ (empirical result): The concept C and having blue sensations have the same extension in the real world.

Premiss P2′ (conceptual insight): A sentient being in counterfactual circumstances falls under the concept of having blue sensations iff it is in the same physical-functional state as those sentient beings falling under the concept of having blue sensations in the real world.

Therefore:

(1′) A sentient being in counterfactual circumstances falls under the concept of having blue sensations iff it has the property expressed by C. (From P1′ and P2′.)

Therefore:

(2′) The concept of having blue sensations and the concept C are necessarily coextensional. (From (2′), presupposing that something in counterfactual circumstances falls under the concept C just in case it has the property expressed by C.)

But, according to the claims about the special status of phenomenal concepts defended above, premiss P2′ is *not* a conceptual insight. So there is no parallel argument using an empirical and a conceptual premiss leading to the result that having blue sensations is to have the property expressed by C.

Could there be some *other* argument leading to the same result? Maybe premiss P2′ is *not* a conceptual insight but can be justified in some other way. A possible strategy for the identity theorist is to claim that premiss P2′ is *not* conceptually true but *should* be accepted for certain theoretical reasons.[13] But this strategy is unsuccessful: concepts are characterized by the corresponding essentiality assumptions that are incorporated into the two-dimensional function associated with the concept. The normative claim that we should accept premiss P2 therefore amounts to the normative claim that we *should* use different concepts of our phenomenal properties and that we *should cease* to think about them in terms of phenomenal concepts. But we simply *have* phenomenal concepts and it is hard to even understand what it would be for us to cease to use phenomenal concepts in our thoughts about phenomenal properties.

If the normative strategy is unsuccessful, maybe then the identity theorist can find some *other* way to justify premiss P2′? At this point the dualist may try to show that there is no possible argument for P2′ that does not beg the question against the dualist. Or, of course, he or she may try to find an *independent* argument against the result

[13] For a similar point and critical remarks about the normative claim see Brie Gertler (1999).

(2′) and thus show in a general manner that there cannot be any valid argument for the second premiss.[14]

References

Chalmers, David (1996). *The Conscious Mind. In Search of a Fundamental Theory*, Oxford: Oxford University Press.

____(2002). "Content and Epistemology of Phenomenal Belief", in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Essays*, Oxford: Oxford University Press.

Flanagan, O. (1992). *Consciousness Reconsidered*, Cambridge, Mass.: MIT Press.

Gertler, Brie (1999). "A Defence of the Knowledge Argument", *Philosophical Studies* 93: 317–36.

Hardin, Clyde Hardin (1987). "Qualia and Materialism: Closing the Explanatory Gap", *Philosophy and Phenomenological Research* 48, 281–98.

____(1992). "Physiology, Phenomenology, and Spinoza's True Colors", in A. Beckermann, J. Kim, and H. Flohr (eds.), *Emergence or Reduction—Essays on the Prospects of Nonreductive Physicalism*. Berlin: Walter de Gruyter, 201–19.

Jackson, Frank (1982). "Epiphenomenal Qualia", *Philosophical Quarterly* 32: 127–36.

Loar, B. (1997). "Phenomenal States", in Ned Block, Owen Flanagan, and Guven Guzeldere (eds.), *The Nature of Consciousness*. Cambridge, Mass.: MIT Press/Bradford Books.

Nida-Rümelin, Martine (1996). "What Mary couldn't know", in Thomas Metzinger (ed.), *Phenomenal Consciousness*, Paderborn: Schoeningh.

____(1998). "On Belief About Experiences: An Epistemological Distinction Applied to the Knowledge Argument, *Philosophy and Phenomenological Research* 58 (1): 51–73.

____(1999). "Pseudonormal Vision and Color Qualia", in S. Hameroff, A. Kaszniak, and D. Chalmers (eds.), *Towards a Science of Consciousness III*, Cambridge, Mass.: MIT Press, 75–84.

____(2003). "Phänomenale Begriff", in Ulrike Haas-Spohn (Hrsg.), *Intentionalität zwischen Subjektivität und Weltbezug*, Paderborn: Mentis.

____(2004). "Phenomenal Essentialism", in Ralph Schumacher (ed.), *Perception and Reality*, Paderborn: Mentis.

____(2006). "Grasping Phenomenal Properties", in Torin Alter and Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge*, Oxford: Oxford University Press.

Strawson, Galen (1999). "Realist Materialist Monism", in S. Hameroff, A. Kaszniak, and D. Chalmers (eds.), *Towards a Science of Consciousness III,* Cambridge, Mass.: MIT Press, 23–32.

White, Stephen (2005). "Why the property dualism argument won't go away". Manuscript. Available at http://www.nyu.edu/gsas/dept/philo/courses//consciousness/papers/WHYPDAW.html

[14] An argument of the latter kind that uses some of the ideas presented in this paper is developed in my paper (2004); a revised version will be included in my paper (2006).

9

Rationalism, Morality, and Two Dimensions

Christopher Peacocke

Basic moral principles are known to us a priori. I will be arguing for this claim, trying to say what it means, and discussing its ramifications.

The claim that basic moral principles are a priori was emphasized by Leibniz and, on some natural readings of the texts, endorsed by Kant.[1] Even a self-proclaimed empiricist like Locke sometimes veered towards endorsing this claim of a priori status.[2] Yet the character of this a priori status, and its significance for the epistemology and metaphysics of moral claims, have both been very largely lost in recent discussions of moral thought. I will be arguing that the nature of this a priori status is incompatible with subjectivist, judgement-dependent and mind-dependent treatments of moral thought. Part of the task in establishing this incompatibility is to articulate more precisely the kind of a priori status that is in question here. It is easy to underestimate the problem for mind-dependent theories of moral thought if one starts by understating the sense in which basic moral principles are a priori.

If basic moral principles are a priori in a way that is incompatible with mind-dependent treatments, various tasks become pressing. One task is to develop a conception of the metaphysics and epistemology of morals that respects this status. Another is to address some of the motivations that have made mind-dependent views

I thank Richard Boyd, Alex Byrne, Paul Boghossian, Kit Fine, Christine Korsgaard, Wolfgang Künne, Derek Parfit, James Pryor, Peter Railton, Stephen Schiffer, Tim Scanlon, Pekka Väyrynen, Ralph Wedgwood, David Wiggins and Aaron Zimmerman for valuable discussion and comments. This material was presented in 2000–1 at Birkbeck College London, at the fourth meeting of the Gesellschaft für Analytische Philosophie in Bielefeld, at Cornell and New York Universities, at discussion groups in Oxford, and as the first of my Whitehead Lectures at Harvard University. This paper is an expanded version of my article 'Moral Rationalism', which appeared in *The Journal of Philosophy* 101 (2004), 499–526.

[1] Leibniz, *New Essays on Human Understanding*, esp. Book I, chapter ii, pp. 91–4 in the edition of P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981); Kant, *Groundwork of the Metaphysics of Morals*, at 4: 408 (pp. 62–3) in *Practical Philosophy*, Cambridge Edition of the Works of Immanuel Kant, trans. and ed. M. Gregor (Cambridge: Cambridge University Press, 1996). Rawls argues that only the procedure of the Categorical-Imperative is a priori for Kant, and that moral principles are reached using it only in the presence of empirical information. See J. Rawls, *Lectures on the History of Moral Philosophy*, ed. B. Herman (Cambridge, Mass.: Harvard University Press, 2000), pp. 247–52. If Rawls's reading is correct, it remains that one of the sources of true moral principles is fundamentally a priori.

[2] *An Essay Concerning Human Understanding,* Book IV, chapter 4, section 7.

of this territory so tempting. Evidently, I am not going to do all this in one paper. But after attempting to make out the case against mind-dependent theories, I will try to outline some possible directions of development; and also to identify something I will call "the Subjectivist Fallacy" which can make mind-dependent views of morality seem more attractive than they really are.

One can pursue these questions about the a priori status of basic moral principles as issues of interest in their own right in the subject of morality and its epistemology. But the questions also have a wider significance. The case of moral thought is of interest as a test case for anyone sympathetic to a more general program of moderate rationalism.[3] Moderate rationalism seeks to explain all cases of a priori knowledge by appeal to the nature of the concepts that feature in contents that are known a priori. For the moderate rationalist, the explanations of a priori knowledge in various domains will not involve the postulation of causal interactions with non-physical or non-mental realms. That is what makes it a moderate rationalism. The explanations will also treat the a priori ways of coming to know as rational, as an exercise of reason. That is what makes the moderate position a form of rationalism. What I have to say in this area can be seen as some first steps towards carrying through the moderate rationalist's program in the special case of moral thought. I hope that some of the considerations I offer will be of more general application, and will help in the development of a moderate rationalism in other areas.

1. The Claim of A Priori Status

Here is a first formulation of the claim of a priori status:

> Every moral principle that we know, or are entitled to accept, is either itself a priori, or it is derivable from known a priori moral principles in conjunction with non-moral propositions which we know.

For an illustration of this Initial Thesis, consider the moral proposition that national high school examinations which assume that candidates have first-hand knowledge of vocabulary needed in snowy climates are unfair to those who live in southern states. That is not itself an a priori principle. No amount of a priori reflection would succeed in excogitating it. The moral proposition does, however, follow from two other truths: from the a priori principle that fair examinations will not include questions requiring background knowledge likely to be absent in one geographical group, together with the empirical, non-moral fact that it rarely snows in the southern states.

The Initial Thesis implies that for any moral proposition we are entitled to accept, there is a similar division: into its a priori moral grounds on the one hand, and its a posteriori non-moral grounds on the other. What the Initial Thesis excludes is the irreducibly a posteriori moral ground. The Initial Thesis is in the spirit of, indeed I

[3] See C. Peacocke, "Explaining the A Priori", in *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke (Oxford: Oxford University Press, 2000).

would say a formulation of, Kant's claim that "all moral philosophy is based entirely upon its pure part".[4]

Why should we believe the Initial Thesis? All sorts of heavy-duty theories—theories of the a priori and theories of morality—might be offered in its support. I shall be touching on, and endorsing, some of them later. But the primary reason for accepting the Initial Thesis is not theoretical at all: it rests on the consideration of examples. Consider your belief that prima facie it is good if the institutions in a society are just; or your belief that prima facie it is wrong to cause avoidable suffering; or that prima facie, legal trials should be governed by fair procedures. These beliefs of yours do not, and do not need to, rely on the contents of your perceptual experiences, or the character of the conscious states you happen to enjoy, in order for you rationally to hold them. Understanding of what justice is, of what suffering is, of what a trial and what fairness is, makes these several beliefs rational without justificational reliance on empirical experience. Experience, as Kant said, may be necessary for the acquisition of these concepts, but that does not mean there cannot be propositions involving them that are a priori. Nor is it clear how empirical experience could rationally undermine these beliefs. Empirical information about extraordinary circumstances might convince us that it would be better on this occasion that a trial not be fair. That would not undermine the proposition that prima facie trials ought to be fair; and it is not clear what could. Take any other moral principle that you are entitled to accept: I suggest that on examination, it will always involve an a priori component, in the sense employed in the Initial Thesis.

The epistemic situation in the case of moral principles seems to me broadly similar to that concerning the status of logic and arithmetic. All sorts of heavy-duty theories—philosophical theories about logic, arithmetic and the a priori—can be offered to support the view that logic and arithmetic are a priori. Those theories may or may not be convincing, but they could not be more convincing than the evidence they attempt to explain, such facts as that we are, apparently, justified in accepting that $2 + 2 = 4$, or that AvB follows from A, without justificational reliance on the content of our perceptual experiences, or other conscious states. In both the moral and the arithmetical and logical cases we must of course be prepared for the possibility that these appearances of a priori status are misleading. Anyone who defends the Initial Thesis must address all sorts of challenges, not all of which I can consider here. All I am emphasizing at this point is that there is strong prima facie support for the Initial Thesis from consideration of examples, in advance of any detailed philosophical theory of how or why the Thesis holds.

The Initial Thesis is neutral on the question of whether every true moral principle could be known by us. People who disagree about that could agree in accepting the Initial Thesis. The Initial Thesis concerns only the cases in which a principle is known, and says something about the existence of a priori ways of coming to know the principle.

This does not make the Initial Thesis a mere de facto claim about the moral principles we happen to know. The reasons for accepting the Initial Thesis go beyond

[4] *Groundwork of the Metaphysic of Morals* at 4: 389 (p. 45 in the Cambridge edition, see note 1).

what is provided by inspection of the particular moral principles we actually accept. I will be offering some general grounds for the Initial Thesis that are not dependent upon the particular moral principles we are currently entitled to accept. There is some plausibility in the further claim that the Initial Thesis, if true at all, is itself a priori. In any case, it has the status of a philosophical, not an empirical, claim.

For those who think that it is begging too many questions to formulate a thesis in a form that presupposes the possibility of moral knowledge, we could frame a version, which may be more comfortable for those doubters, that mentions only entitlement to accept. (I myself doubt that this really is weaker, but I mention it so that we can focus on the essential issues.) Any interesting version of the Initial Thesis must, however, make some use of some distinction between proper and improper acceptance of a moral principle. It could not be formulated in terms of mere acceptance.

The Initial Thesis is cagily formulated using "we". It will not be true of each individual thinker that every moral principle he is entitled to accept is either a priori, or derivable from a priori moral principles and non-moral propositions he knows. Moral knowledge, like any other kind of knowledge, can be acquired by testimony. An empirical moral principle may be so acquired, and when it is, the acquirer himself need not know the a priori grounds of the empirical moral principle he learns through conversation. Nevertheless, someone must know or once have known them if the moral belief he acquires by testimony is to have the status of knowledge. The Initial Thesis is a thesis about actual epistemic grounds, in the epistemic community as a whole over time. The Thesis goes far beyond claims about the mere possibility of grounds.

What do I mean by "a priori"? For an intuitive, overarching characterization of a standard notion, we can say this: a thinker's judgement is a priori if it has an operative justification or an entitlement that is independent of the representational content or kind of the thinker's perceptual experience, and of her other current conscious states. So the judgement "There's a window over there", when the thinker makes it because he sees a window to be over there, is not a priori, because it endorses the content of the thinker's perceptual experience. The judgements "I'm in pain" and "I'm imagining standing on a beach" are not a priori when the thinker's operative justification or entitlement lies in the character of his current conscious states, his pain or his imaginings. In all these cases—of seeing the window, of the pain, and of the imagination—there is a way in which the judgement comes to be made and whose status as justifying or entitling is dependent on one or another features of perceptual experience, or of other conscious states. The way itself is not a priori, we might say. By contrast, judgements to which the thinker is entitled because the thinker, or someone else, has a proof of their contents are a priori by this umbrella criterion.

The umbrella characterization covers two fundamentally different species of the a priori. As I implied in the introductory remarks, it is important to distinguish them, both from each other and from related notions in the territory, if we are to have a clear view of the significance of the senses in which basic moral principles are a priori.

The two species of the a priori can be introduced by first considering a much more general auxiliary notion. This more general notion in its most abstract form stretches far beyond the a priori. It is the notion of a judgement with a given intentional

content being true in any circumstances in which it is reached in a given way. A judgement "I'm in pain" that the thinker makes rationally because she consciously experiences pain falls under this general notion. In any circumstances in which a thinker comes to make the self-ascription of pain by rationally responding to her conscious experience of pain, her self-ascription will be true. A judgement of a logical truth reached by accepting a proof of it equally falls under the same notion. I label this very general notion that of *p*'s being *judgementally valid* with respect to a given way.

It is important that the judgemental validity of a content with respect to a given way turns only on the truth of the content in circumstances in which it is in fact judged (and reached in the given way). In assessing judgemental validity with respect to a given way, we do not have to consider whether the content is true in circumstances in which it is not reached in that way. Nor do we have to consider whether the content has any kind of necessity.

Various famous concepts in philosophy are variants of this core notion of judgemental validity. Descartes was particularly interested in those contents with the following property: that there exists a way with respect to which they are judgementally valid, and which is indubitably so. Descartes' description of something that is "necessarily true whenever it is put forward by me or conceived in my mind" is a variant, with additional restrictions, of the core idea of the judgementally valid.[5]

We can make use of this auxiliary notion of the judgementally valid in distinguishing the two species of the a priori that I want to distinguish. The first notion of the a priori to be distinguished is simply a restriction of the notion of judgemental validity. I say that

> *p* is *judgementally a priori* with respect to a way W just in case it is judgementally valid with respect to W, and the way W is an a priori way.

The judgementally a priori includes some classical self-verifying cases. When the content "I am thinking" is judged, not as a report on the thinker's own recent conscious states, but because the thinker appreciates, on the basis of his grasp of the concepts it contains, that it will be true in any circumstances in which he judges it, the content is judgementally a priori with respect to this way. The same applies to "I hereby judge that water is H_2O". The judgementally a priori will also include such traditionally acknowledged examples of the a priori as contents reached by mathematical proof.

Closely related to the judgementally a priori is a notion which is not itself a form of the a priori. Consider someone who makes the second-order self-ascriptive judgement "I judge that $13 \times 5 = 65$", or makes the judgement "I judge that if A & B, then A". Suppose this thinker comes to make these judgements by the procedure that Gareth Evans described. That is, to quote Evans's description, "I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*".[6] So—to take the arithmetical

[5] Second Meditation, in *Philosophical Writings of Descartes: Volume II*, trans. J. Cottingham, R. Stoothof, and D. Murdoch (Cambridge: Cambridge University Press, 1984), p. 17 (at p. 25).

[6] *The Varieties of Reference* (Oxford University Press, 1982), section 7.4, p. 225.

case—our thinker makes the self-ascription as follows. She first considers the first-order question of whether 13×5 is in fact 65. This will involve an arithmetical computation. In reaching the conclusion that $13 \times 5 = 65$, the thinker will not (or need not) be relying on a justification or an entitlement that involves the content or character of her conscious states. Employing the procedure described by Evans, our thinker then moves from her conclusion that $13 \times 5 = 65$ to the self-ascriptive judgement "Yes, I believe that 13×5 is 65". This self-ascriptive content is judgementally valid with respect to this method of coming to judge the self-ascriptive content.

The earlier stages of the way employed in reaching this self-ascription are a priori. The example can be such that only a priori premises, and transitions, are used by the thinker in reaching the conclusion that $13 \times 5 = 65$. But is the whole method of coming to make the self-ascription also itself a priori? It is not. The transition the thinker makes from judging that $13 \times 5 = 65$ to judging that she judges that $13 \times 5 = 65$ is one to which the thinker is entitled only because she actually judges that $13 \times 5 = 65$. It is not like a case of perceiving a proof, in which the thinker has access to something which gives a rationale for the conclusion independently of the thinker's perception of the proof. In this self-ascriptive case, the thinker's making the first-order judgement is part of the rationale for the self-ascription.

In Evans's procedure for self-ascription, it seems, as Evans emphasizes, that the thinker does not engage in introspection in self-ascribing beliefs to herself. It follows from what we have just said that we cannot capture the respect in which the procedure is not introspective by saying that contents known by it are judgementally a priori: for they are not. It follows that Evans's point must be elucidated some other way. We can introduce the notion of a way of coming to judge a content being *non-introspective*, as follows. Such a way of coming to make a self-ascription is non-introspective in case: (a) other than the final judgement reached in employing the way, the contents employed in using it are not about the thinker's own mental states or events; and (b) the means by which the thinker comes to accept these contents in employing the way do not involve checking on his own mental states or events. Evans's procedure, as used in self-ascribing the belief that $13 \times 5 = 65$, meets this condition for being a non-introspective way. So too does the analogue of Evans's procedure for self-ascribing visual perception. One way of coming to judge, and to know, "I see that there's a desk in front of me" is to investigate the world around one by looking, and making that self-ascription just in case one sees that there is a desk in front of one. In neither this case, nor in the arithmetical case, is the self-ascription reached by an a priori way. But in both cases, the self-ascription is reached by a way that is non-introspective.

Other attitudes besides acceptance or judgement can be self-ascribed by ways which are non-introspective. Consider a self-ascription of an intention "I intend to answer the objection in the next sentence". A thinker may come to make the self-ascription in the following way. She goes through the procedure of deciding what to write next, and then makes the self-ascription of the intention if and only if she decides to answer the objection in the next sentence. This is the analogue for self-ascriptions of intention of Evans's procedure for the self-ascription of belief. The thinker's reasons for making her decision may have to do with such matters as what is a good reason for

thinking what, and what layout of the argument will reflect this. In coming to make the self-ascription of the intention in this fashion, the thinker will not be checking on her own mental states or events. The way in which the self-ascription is reached is non-introspective. The way is also not a priori. The thinker's entitlement to make the self-ascription of the intention depends upon her actually making the decision in question.

So much by illustration of the notion of the judgementally a priori, and its difference from what is judgementally valid in a non-introspective way. The second notion of the a priori of which I will be making extensive use I call the "contentually a priori":

> *p* is *contentually a priori* with respect to a way W if W is an a priori way of coming to know *p*, and W is also a way that ensures the following: the content *p* of the judgement it yields is true in the actual world, whichever world is labelled as the actual world, and is true regardless of whether that way W is used, and of whether the conditions of its use are met, in the world that is labelled as the actual world. [7]

Here the phrase "whichever world is labelled as the actual world" is not meant to mean "I don't care what the actual world is like". "*p* is true in the actual world, whichever is labelled as the actual world" here means: for any possible world, if it were actual, *p* would be true when evaluated with respect to it.

To say that something comes to be known in a way that ensures that it is true in the actual world, whichever is the actual world, is not to say this: that someone who comes to know something in this way thereby comes to know *that* it is true in the actual world, whichever is the actual world. The situation is quite parallel to the more straightforward case of the intuitive notion of an a priori way of coming to know some content. A person's entitlement can be a priori without her exercising, or even possessing, the concept of the a priori. The same point applies to the contentually a priori. A person can come to know something that is contentually a priori with respect to the way in which she comes to accept it, without herself exercising or even possessing the concept of the contentually a priori. The fact, however, that there is a way of coming to accept a given content that does ensure that it is true in the actual world, whichever is the actual world, is something striking, and in need of philosophical explanation.

Being contentually a priori is a relation, between a content and a way. It will often be convenient to use an existential quantification of the relation. We say that something is contentually a priori tout court if there is some way with respect to which it is contentually a priori.

Those who are not made queasy by the whole idea of the a priori would count amongst the contentually a priori propositions the following: the known logical truths; known arithmetical truths; and propositions such as "If I exist, and this place

[7] This is a more refined version of the distinction drawn between "the judgementally a priori" and "the contentually a priori" in C. Peacocke, "Implicit Conceptions, the A Priori, and the Identity of Concepts", in *Concepts*, ed. E. Villaneuva, Vol. **9** (1998) of *Philosophical Issues* **9** (1998) (Atascadero, Ca.: Ridgeview), 121–41, at p. 136.

here exists, then I am here", "No shade is both a shade of red and a shade of green", and "If *p*, then Actually *p*". As some of these examples illustrate, and as the writings of Kripke and Kaplan made clear, something can be contentually a priori without being metaphysically necessary.

In modal semantics, Martin Davies and Lloyd Humberstone very helpfully introduced an operator "Fixedly". Its semantical clause states that "Fixedly *p*" holds at a given world in a given model just in case it holds in that world in any model differing only in which world is labelled as the actual world.[8] All the contentually a priori propositions I just mentioned hold Fixedly Actually in the sense of Davies and Humberstone. That is, if we preface them with the pair of operators "Fixedly" and "Actually", in that order, the result is true.

Enthusiasts for philosophically significant formal semantics will also be struck by the affinity between the contentually a priori and David Kaplan's notion of validity in the logic of demonstratives, that is, the notion of truth with respect to every context in every structure.[9] The two notions are cousins. Kaplan is concerned with language rather than thought. For his semantical purposes, Kaplan does not need to be concerned with ways of coming to know. But the property of being contentually a priori and the property of being valid in Kaplan's logic of demonstratives are not distant cousins. For a sentence-type to be true in a given context, in Kaplan's treatment, it is not required that there exist an utterance of that sentence in that context.[10] In Kaplan's development, an expression-type can be true with respect to a context without being uttered in that context. So, unlike "I am here", "I am uttering something" would not be valid in the logic of demonstratives (if the object-language were extended to include "utter"). Kaplan's notion of validity in the semantics of demonstratives is therefore not a variant of "true whenever uttered". It is closer to being a linguistic analogue of the contentually a priori than of the judgementally a priori.[11]

There is a sharp difference in extension between the judgementally a priori and the contentually a priori. Not everything that is judgementally a priori is contentually a priori. Simply considering the matter in the abstract, one should expect this. For a content to be judgementally a priori it is required only that it be true in each world in which it comes, by a certain route, to be judged: whereas to be contentually a priori a content must be true in the actual world, whichever is the actual world, regardless of whether it is judged, or how it comes to be judged. (In the reverse direction, anything that is contentually a priori is judgementally a priori. If something is true in the actual world, whichever is the actual world, it will be true when evaluated with respect any world in which it is judged. It also seems that if a way of coming to judge something yields a content which is true in the actual world, whichever is the actual

[8] "Two Notions of Necessity", *Philosophical Studies* 38 (1980), 1–30, p. 2.

[9] "Demonstratives" in *Themes from Kaplan* (New York: Oxford University Press, 1989), ed. J. Almog, J. Perry, and H. Wettstein, second definition on p. 547.

[10] "But to develop a logic of demonstratives it seems most natural to be able to evaluate several premises and a conclusion all in the same context. Thus the notion of *φ being true in c and A* does not require an utterance of φ." ("Demonstratives", p. 546).

[11] Similarly, all the distinctive examples discussed by Kaplan of sentences which are valid in the logic of demonstratives are ones whose intentional contents are contentually a priori.

world, it cannot be justificationally dependent on the content or kind of experience or conscious states a thinker enjoys.)

The examples bear out the expectation of a difference in extension. Some self-verifying judgements are judgementally a priori, but they do not have the property of being true in the actual world, whichever is the actual world. Worlds in which I am not thinking now, or not judging that water is H_2O, could have been actual.

The contentually a priori is also different in extension from being judgementally valid with respect to a non-introspective way. Worlds in which I never consider whether $13 \times 5 = 65$, or in which I consider the matter, but make calculating errors, could have been actual.

Consider a way W which, when used, leads to judgement of a content that is judgementally a priori but not contentually a priori—that is, it leads to something which is *merely* judgementally a priori, as I will say. The explanation of why such a way W leads to a true judgement has to mention that fact that certain contents are actually accepted, or stand in other psychological relations, when the judgement is reached in that way. This applies to the explanation of the truth of such self-verifying judgements as "I am thinking" and "I (hereby) judge that water is H_2O". The explanation of why their contents are true must mention the fact that the judgements are actually made.

A slightly different, but analogous, point holds for judgements reached by Evans's procedure for self-ascription. When employed in coming to judge "I judge that $13 \times 5 = 65$", Evans's method yields knowledge only because in executing that procedure, and reaching this result, the subject also comes to accept that $13 \times 5 = 65$. The same applies to the procedure we mentioned for self-ascribing an intention. The procedure works only because it involves the formation of a certain attitude in its execution—in the example, it was the decision to answer the objection in the next sentence.

All these cases contrast with acceptance of the first-order content "$13 \times 5 = 65$" on the basis of an arithmetical computation. The computational method is guaranteed to yield a result that is true in the actual world, whichever is the actual world, without reference to anything involving acceptance of the intermediate stages, or indeed anything psychological at all. That is why the first-order judgement of $13 \times 5 = 65$ meets the stronger condition of being contentually a priori.

With the distinction between the contentually a priori and the judgementally a priori in hand, we can return to the Initial Thesis. At first blush, moral principles that are a priori do not seem to be merely judgementally a priori. They do not seem to be true only in worlds in which they come to be judged in a certain way. First blushes can be misleading, and have superficial causes, and I will return to the issue. For now, I want to propose, consider and defend the Initial Thesis in a sharpened and strengthened form, in which it concerns the contentually a priori. The Sharpened Thesis states:

> Every moral principle that we know, or that we are entitled to accept, is either contentually a priori, or follows from contentually a priori moral principles that are known in conjunction with non-moral propositions that we also know.

This needs argument and defence against a variety of challenges. I will try to provide some of what is needed a few paragraphs hence. First I offer some observations intended to bring out the nature of this Sharpened Thesis.

The Sharpened Thesis corresponds closely to parallel theses in two other areas in which knowable truth seems to be truth that is, at a fundamental level, contentually a priori.

The first of these areas is that of metaphysical necessity, whose partial parallels with the moral case I will consider at several points. Each truth that contains a metaphysical modality, and that is also known to us, seems to be either itself contentually a priori, or it seems to follow from truths each of which is either a modal contentually a priori truth, or is an a posteriori non-modal truth. It is necessary that Tully is Cicero. That modal truth is a posteriori. But it is a consequence of an a priori modal truth—the necessity of identity—together with the a posteriori but also non-modal truth that Tully is Cicero. It has become a familiar claim about metaphysical necessity that every modal truth has its source in principles which are either necessary and a priori, or non-modal and a posteriori.[12] What is excluded is an irreducibly a posteriori modal truth. The a priori modal principles that are fundamental under this conception of metaphysical necessity are plausibly contentually a priori, and not merely judgementally a priori.

The second case paralleling the Sharpened Thesis is that of evidential and confirmation relations. Many instances of evidential and confirmation relations are a posteriori. But it is arguable that each of them has an a priori component. A certain kind of rash confirms that an illness is meningitis. That is certainly a posteriori. But it rests on the a priori principle that a suitable range of instances gives non-conclusive support for a generalization, together with the truths about the presence of the rash in previous instances only in cases of meningitis, truths that are not themselves about the confirmation relation. What is excluded is an irreducibly a posteriori truth essentially about confirmation. Again, the notion of the a priori on which these are plausible claims is that of the contentually a priori.

Since evidential and confirmation relations are normative relations, this second case does more than merely provide a parallel example. It further suggests a general hypothesis: that there is a significant range of normative kinds, such that each truth of that kind has an a priori component. This thought will be resurfacing at several points later on.

The Sharpened Thesis has a more general epistemological feature. There has in discussions of justification and the a priori long been circulating an argument to the effect that in any domain in which justifications and reasons exist, some reason-giving relations must have an a priori status.[13] (Here we have the general hypothesis that all normative truths have an a priori component resurfacing already.) It is hard to

[12] See C. Peacocke, *Being Known* (Oxford: Oxford University Press, 1999), chapter 4, "Necessity".

[13] For an overview and one kind of defence, see L. BonJour, *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1988), p. 5 ff. For a somewhat different argument for the view that all instances of the entitlement relation are fundamentally a priori, see C. Peacocke, "Three Principles of Rationalism", *European Journal of Philosophy* 10 (2002), 375–97, esp. Principle III, the Generalized

see how justification and the making of judgements for good reasons could ever get started if all reason-giving relations were a posteriori. My own view is that this traditional argument is sound, when it is properly framed. There are all sorts of ways of mishandling the idea, some of which have to do with certainty. One such way of mishandling the idea is the view that if anything is probable, something must be certain.[14] But the idea that justification or entitlement could not get started unless some principles or relations are a priori can be developed without any commitment to the existence of such certainties.

If the reasoning of the traditional argument is sound, it applies as much in the domain of moral thought as it does in the area of empirical thought. Our Sharpened Thesis that all moral principles we are entitled to accept have a contentually a priori component dovetails with the traditional argument about justification. The Sharpened Thesis alludes to what must exist within the moral domain if the traditional argument is sound.

The Sharpened Thesis also has metaphysical ramifications, but I will first attempt to understand and explain its epistemic aspects.

2. The Claim Defended

A first objection to the claim of a priori status for basic moral principles may be that a thinker's impression, perhaps after some reflection, that a moral principle is correct is something that plays both a causal and a rational role in the thinker's acceptance of the moral principle. Why then is this impression not a conscious state whose role implies that basic moral principles are not a priori after all?

There must be something wrong with this objection, because such conscious states, playing a causal and a rational role, are present in clear cases of a priori status. A thinker may reflect rationally, and after her reflection, be left with the impression that a principle is a logical law. The thinker's impression will be both causally and rationally operative in her acceptance of the principle as a law. It is rational, in the absence of reasons for doubt, to accept the outcome of such processes of reflective thinking. This can be an a priori way of coming to know the law.

What more specifically is wrong with the objection is that in the examples in question, the impression is not a justification. The impression of correctness is itself a rational response to conditions that give grounds for thinking that (say) gratuitous infliction of pain is prima facie wrong, or give reasons for thinking that the logical law is valid. In the former case, the fact that pain is subjectively awful provides such grounds; in the logical case, the justifying condition for a reflective thinker

Rationalist Thesis. If it is true that all entitlement is fundamentally a priori, that is not in itself alone an argument for moral rationalism. There can be entitlements in domains for which forms of subjectivism or mind-dependence hold. In those cases, the entitlements are dependent upon certain mental conditions holding. My position is that there is no such dependence in our entitlement to hold moral propositions.

[14] A theme in C. I. Lewis's writings. See *Mind and the World Order: Outline of a Theory of Knowledge* (New York: Dover, 1956), for instance pp. 311–12.

must include the fact that the law is true under all relevant assignments, or can be derived from such laws. The thinker has an impression of correctness only because he appreciates these justifications. Since the impressions of correctness in these examples are not themselves justifications, they cannot be used to support the claim that the thinker's operative justification in the moral or the logical cases is the character of one of his mental states.

This point is entirely consistent with the impression playing a causal role in the rational process leading up to the thinker's acceptance of the content. Of course the thinker would not have made the judgement in question if he had not had the impression that the content is correct. But that does not make the impression into a justification.

We can further emphasize the distance between impressions and justifications by considering their relations to correctness. For anything that is a justification for accepting a given content, there must be an account of why that justification entitles the thinker to judge that the content is true—an account of the relation between justification and truth, in short. An explanation of how a judgement comes to be made that includes reference to an impression of correctness is not by itself an explanation of why that method of reaching the judgement is a correct method. For that, we need an account that mentions that to which the impression is a rational response, when it is a rational response.

This treatment still sharply separates the a priori cases from those of perceptual knowledge. Suppose you come to have the perceptual knowledge "That flower is yellow". Your impression that this is a correct content is one to which you are entitled by the character of your perceptual experience; so the judgement is squarely a posteriori, indeed the paradigm case thereof. The explanation of why this is a correct way of reaching a judgement "That flower is yellow" would certainly have to mention the perceptual experience, as a source of non-inferential information about the world. Your impression that the content "That flower is yellow" is correct in these circumstances is parasitic on the justifying or entitling role of the mental state of perceptual experience, with its relation to correctness.

There are some kinds of example in which an impression is itself entitling. Propositional, non-autobiographical memory, of the sort likely instanced by your memory that the Bolshevik Revolution occurred in 1917, provides perhaps the clearest example. Here your impression is not a rational response to anything. Such impressions are entitling if we are entitled to take the deliverances of a memory faculty at face value. But the status of propositional memory that *p* as entitling depends on the thinker's entitlement to accept the information that *p* when he originally acquired it. At the time of acquisition, what is entitling cannot be your memory impression. It may be an impression of correctness you have at the time of acquisition because some honest, knowledgeable interlocutor has informed you that *p*. This too can be entitling: but again, it seems that its status as such traces back eventually to the acquisition of an entitled belief that *p* where the entitlement does not consist solely in an impression of correctness.

In the logical case, the reasons producing the impression that it is correct that a certain principle is a logical truth are reasons that are experience-independent and

mind-independent. They consist in the existence of a proof that there is no falsifying assignment. As always, we must distinguish between the proof itself and access to the proof. Access to the proof must involve psychological matters: but that does not make what is accessed, the proof itself, into something mind-dependent.

Some theories treat the impression of the correctness of a moral principle as something which is not the appreciation of a reason which is explicable independently of the thinker's reactions on thinking about the principle, or its instances. There is a large subclass of such theories that treat moral properties as mind-dependent. Many different varieties of theory involve such mind-dependence. It is present in Christine Korsgaard's idea that the source of normativity is an agent's endorsement of "a certain way she looks at herself, a description under which she finds her life worth living and her actions worth undertaking".[15] It is present in judgement-dependent theories, in various forms of subjectivism, and in a range of dispositional theories, where the dispositions in question concern mental properties.[16]

Mind-dependence also seems to me to be present in Simon Blackburn's treatment of moral thought, even though he himself explicitly denies that his view involves mind-dependence.[17] Blackburn describes his view as a form of expressivism: in making moral judgements, one expresses certain mental states, which, he holds, can be characterized as non-representational. Blackburn writes of an earlier paper of his "I said that the moral proposition was a 'propositional reflection' of states that are first understood in other terms than that they represent anything, and that remains the core claim" (p. 77). A distinctive feature of Blackburn's position is that he allows that moral propositions can be assessed as true or false, and he appeals to a minimalism about truth in support of his position (p. 79). It is, however, very hard to see how it can be denied that, under his approach, the conditions under which someone is correct in asserting a moral proposition have something to do with expressed mental states. To make this point is not to say that it is a consequence of Blackburn's position that someone making a moral claim is saying something about mental states. It is not a consequence. But it is equally not a consequence of a classical secondary-quality view of the property of being red that someone who says that an object is red is saying something about the experiences produced by that object. The classical secondary-quality view does, nevertheless, treat colour properties as mind-dependent. In neither the moral nor the colour case should the philosophical theory be put into

[15] *The Sources of Normativity*, ed. O. O'Neill (Cambridge: Cambridge University Press, 1996), p. 249.

[16] For discussion of such approaches, see C. Wright, *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press, 1992), ch. 3, Appendix to ch. 3, and ch. 5; D. Wiggins, "A Sensible Subjectivism?", in his *Needs, Values, Truth* (Oxford: Blackwell, 1987); and the papers by D. Lewis, M. Johnston, and M. Smith in the Symposium "Dispositional Theories of Value", in *Proceedings of the Aristotelian Society, Supplementary Volume* LXIII (1989), 89–174.

[17] Blackburn has developed his view over many years. For a recent overview of his position, see his *Ruling Passions: A Theory of Practical Reason* (Oxford: Oxford University Press, 1998). For a statement of his view that his quasi-realism does not commit him to mind-dependence, see for instance his answer to question 2, pp. 311–12 of *Ruling Passions*. Page references appended to quotes from Blackburn are to this book. Blackburn no longer describes his view as 'projectivism', because it makes it sound as if projecting attitudes involves some kind of mistake: see p. 77.

the content of what the person is saying. It remains the case that, on both theories, the philosophically fundamental account of what it is for an utterance of a moral, or a colour, predication to be correct has to make reference to mental states.

For a theorist who holds that there is no such thing as a moral proposition, the question of correctness would not even arise, and such a theorist might reasonably rebut the ascription to him of the view that the correctness of moral propositions is mind-dependent. But that is not Blackburn's position. I discuss Blackburn's detailed reasons for rejecting the ascription to him of a mind-dependent treatment of morality later in this section.

I now want to raise the following question: can theories which treat the correctness of moral proposition as mind-dependent explain the apparent fact that basic moral principles are contentually a priori?

To separate the issues clearly, I first consider what the mind-dependent theorist can explain. Suppose we have some specific form of mind-dependent approach to moral norms. Suppose too that a thinker judges in ways acknowledged by that theory as suitably sensitive to the mind-dependent properties that he says are constitutive of moral norms. It will then hold according to that theory that the moral principles so reached will be true in any circumstances in which they are so reached. That is, under this mind-dependent theorist's conception, there is a way of reaching moral contents with respect to which they are judgementally valid.

It is hard to see how they could also be judgementally a priori. Under a mind-dependent treatment, the entitlement to make the moral judgements is constitutively dependent upon the instantiation of the mind-dependent properties to which the moral judgements are sensitive, when the thinker is judging knowledgeably.

Can the mind-dependent theorist provide for non-introspective ways of coming to know moral propositions? It seems to me that he can allow for that. If statements of a certain kind are regarded as having mind-dependent truth-conditions, it does not follow that coming to know the truth of such a statement must (even in basic cases) involve checking on the thinker's own current mental states, or on anyone else's mental states. Statements about belief are certainly mind-dependent. Evans's procedure for self-ascription shows that a fundamental procedure for self-ascription may nevertheless involve looking outwards towards the world, not inwards to one's own mental states. Take the mental states to which, according to the mind-dependent theorist, a thinker must be sensitive if he is to be making moral judgements knowledgeably. If those mental states are not themselves about other mental states, the mind-dependent theorist can, it seems, consistently embrace the existence of non-introspective methods of coming to judge, knowledgeably, that certain moral propositions hold. Moral emotions, for example, are directed outwards to events, states of affairs and other people, and are not at all well-described as in general involving introspection of one's own mental states. While there are many good questions about whether the mind-dependent theorist can properly characterize the mental states in terms of which he wants to explain moral thought, I think we can still grant the conditional that if, within the terms of his own theory, he has access to those mental states we normally express in our moral thought, he can legitimately claim that the ways of coming to know which he endorses are non-introspective.

Still, this is not to say that they are a priori. In particular, it does nothing to show that basic moral judgements are contentually a priori. The challenge to the mind-dependent theorist is to answer these questions: must not his theory imply that were our morality-generating sentiments to be different, what is actually wrong would no longer be so? If it does have that implication, he cannot explain the fact that basic moral principles are true in the actual world, whichever is the actual world, since it seems that they would not be true if the actual world were one in which we had different morality-generating attitudes. Can the mind-dependent theorist show that basic moral principles are true in the actual world whichever is the actual world?[18]

This crucial question has multiple readings. On a theory according to which psychological states are, in one way or another, the source of norms, in order to articulate this question more precisely, we need to introduce some double-indexing. We need to use the notion

> proposition P, when evaluated from the standpoint of psychological states in world w1, holds with respect to world w2.

We can abbreviate this to P(w1, w2). It cannot be begging any questions against mind-dependent treatments to employ this doubly-indexed notion. The first parameter makes explicit the dependence that the mind-dependent theorist himself needs to use in articulating his own theory. The second parameter is just assigned whatever world is the one with respect to which the proposition P is being evaluated. So in the case of a mind-dependent theory of morality in particular, "P(w1, w2)" means that proposition P, when assessed according to the moral standards said to result from thinker's psychological states in world w1, holds with respect to w2.

Here it helps to draw up some matrices, analogous to those introduced by Robert Stalnaker.[19] The mind-dependent theorist of moral thought is committed to holding that in each world, there is some set of basic attitudes in terms of which moral truth, or entitlement to moral judgement, is elucidated philosophically and on which the correctness of moral claims depends. We can use the notation "Atts_i" for such postulated basic attitudes as are held by thinkers in world *i*. Each matrix corresponds to a given moral statement S (as we can neutrally put it). In each column of the matrix, we hold constant a parameter of the form Atts_i for some fixed world *i*. The various entries in the column specify the truth-value of the statement S at a given world, with respect to the constant parameter Atts_i. So suppose that under the basic attitudes of world *i*, an action of type A is prima facie good (in some given respect). We can suppose that this is a basic evaluation, and not subject to empirical variation, under the given standards. So in the column for Atts_i, every entry is a "T" for true. But under the

[18] Since I already argued that being contentually a priori implies being judgementally a priori, and suggested that under the mind-dependent view, basic moral principles are not judgementally a priori, I already have an argument that the mind-dependent theorist cannot explain why basic moral principles are contentually a priori. But the arguments of particular theorists, such as those of Blackburn considered below, to the effect that there is no problem here, mean that we have to consider the case of the contentually a priori separately.

[19] *Context and Content* (Oxford: Oxford University Press, 1999), Introduction and chapter 4, "Assertion".

different attitudes of worlds *j* and *k*, such an action-type is not prima facie good; again we suppose that this is a basic evaluation. So the matrix for the statement "Actions of type A are prima facie good" might be as follows:

	$Atts_i$	$Atts_j$	$Atts_k$
i	T	F	F
j	T	F	F
k	T	F	F

Our question was whether the mind-dependent theorist could explain the contentually a priori character of basic moral principles, that is, could explain the fact that they are true in the actual world, whichever is the actual world. There are clearly several possible readings of the phrase "true in the actual world, whichever is the actual world" when we have double indexing.

We can distinguish three features which may be present independently of one another when a reading of this phrase is formulated in terms of the P(w1,w2) notation. The three features correspond to positive answers to these three questions:

(1): Is the first place, occupied by "w1" in "P(w1,w2)" universally quantified? In the matrix notation, this is equivalent to the question: are we considering a condition that universally quantifies over columns? In terms of the substantive philosophy, this first question is asking: are we considering variation in respect of the postulated basic attitudes?

(2): Is the second place, occupied by "w2" in "P(w1,w2)", universally quantified? In matrix notation: are we speaking of what holds in all entries in any given column? In terms of the substantive philosophy: are we considering variation of possible worlds as points of evaluation, with respect to a given set of basic attitudes?

(3): Are the variables or terms of the relation identified? That is, are we concerned with some condition concerning the instantiation of some monadic property involving the proposition P, of the form λw [. . . P. . . (w, w)]? In matrix notation, are we concerned with what holds along the diagonal?

Bearing these distinctions in mind, we can then distinguish at least the following readings of "P is true in the actual world, whichever is the actual world". (I prescind from relabellings of worlds as the actual world, so as not to distract attention from the central point):

Reading (A): For any world w, P(w, w).

This is equivalent to having the entry True at each cell on the diagonal of the matrix that runs from top left to lower right. We can call this "the diagonal reading". It means this: take any world, and the alleged basic morality-generating attitudes of that world, the proposition P will hold in that same world. On this reading, we have both identification of variables and universal quantification of the monadic property λw[P(w, w)].

Reading (B): For any world w, P(@, w).

Here there is no identification of variables, and universal quantification only with respect to the second place. This we can call "the vertical reading", since it fixes

on Atts@, and for this reading to hold all the entries in the column Atts@ must be "True".[20] It means this: take our alleged morality-generating attitudes, and hold them fixed: then P holds in every world, when evaluated with respect to those attitudes so held constant.

There are variants of the diagonal and the vertical readings when one considers interactions with the labelling of a world as the actual world; but I will concentrate just on the diagonal reading (A) and the vertical reading (B).

In his quasi-realist writings, Blackburn seeks to address the natural objection that even if our evaluative attitudes were different, that would not make actions which we actually hold to be wrong into morally acceptable actions. Here is how he replies (I take a recent passage, which gives an answer that he has also developed in several other places):

> According to me, 'moral truths are mind-dependent' can *only* summarize a list like 'If there were no people (or people with different attitudes) then X . . .', where the dots are filled in with some moral claim about X. One can then only assess things on this list by contemplating the nearest possible world in which there are no people or people with different attitudes but X occurs. And then one gives a moral verdict on that situation.[21]

Here Blackburn is following broadly the structure of interpretation (B), the vertical reading. The "moral verdict" of which he speaks is reached by employing our actual standards, which is why he holds that the objection fails.

In a similar spirit, one might imagine a defender of Blackburn's position saying that, on his position, basic moral principles are indeed contentually a priori because we apply our actual basic moral standards, and if we do so, then whichever world is the actual world, the basic moral principles will be correct with respect to it.

Here I protest. On the quasi-realist's theory, the acceptability of basic moral principles depends on some psychological attitudes. However this dependence is formulated, it must be possible to consider which propositions are correct when we vary the standpoint of evaluation, that is, when we vary the first parameter, as in (A). Take a specific moral principle identified by its content, say "Prima facie the infliction of avoidable pain is wrong". Now consider the claim

> For any world w, prima facie the infliction of avoidable pain is wrong (w, w).

It seems to me that the quasi-realist, like other mind-dependent theorists, must say this is false. It is false at those entries in the diagonal for worlds in which we have different attitudes to the infliction of avoidable pain. The mind-dependent theorist has

[20] There will also be a generalized perpendicular reading, which asserts of an arbitrary world considered as actual what the preceding reading asserts only of the actual world. So the generalized perpendicular reading asserts

> (C) For any world w considered as actual, and for any world u, P(w/@, u).

One can also formally distinguish the cases ∀w P(w,@) and ∀w∀u P(w,u).

[21] *Ruling Passions*, p. 311.

not, by his own lights, excluded those worlds. Unless the quasi-realist, or more generally any other mind-dependent theorist, has some way of showing that our basic evaluations could not have been different, I do not see how the mind-dependent theorist can avoid a commitment to denying this most recently displayed claim. In short: the objection to mind-dependent views concerns the diagonal reading, and the objection is that the mind-dependent theorist has not explained, by his lights, why there cannot be an entry "False" somewhere on the diagonal. It cannot be an adequate answer to this objection that there are no "False" entries on the vertical that corresponds to the actual world. Correspondingly, Blackburn's defence cannot show that moral principles are contentually a priori in the sense of the diagonal reading, interpretation (A).

It may be helpful in clarifying the distinction between the diagonal and the vertical readings to fix on some very simple concepts where we would also want to invoke double-indexing. It seems to be widely agreed that things would not stop being red if humans lost their colour vision, and saw only in shades of grey. It is entirely consistent with this point to hold that which colours things have is in some way constitutively dependent upon how humans actually perceive them (in circumstances in which they have not lost their colour vision). If one does hold that further claim, the right way to formulate the dependence is not in terms of counterfactuals like "If we were not to see things as red, they would not be red", or any more sophisticated variants thereof. Such counterfactuals are evaluated from the standpoint of how humans perceive things in some central normal cases (that is, evaluated holding fixed the first parameter), and so cannot capture the intended dependence. But it is possible to formulate the proposed dependence all the same, either at a meta-level, or using some analogue of Davies and Humberstone's "Fixedly Actually" operator. The best way of doing this would again depend on the resolution of various auxiliary issues, but one simple formulation of the suggested dependence is this:

> There is no physically individuated property Q such that it is Fixedly Actually the case that objects with Q are red.[22]

For an arbitrary physical property Q, our imagined mind-dependent theorist will be committed to rejecting the claim that

> (DC) For any world w, Q-objects are red (w, w).

This is precisely parallel to the mind-dependent theorist of morality's commitment to rejecting the claim

> (DM) For any world w, prima facie the infliction of avoidable pain is wrong (w, w).

I am, then, committed to disagreeing with Blackburn's attitude to the seeming meta-level question of whether, on his view, moral truths are mind-dependent. He writes

[22] Though of course this theorist will accept that

For any world w, Q-objects are red (@, w)

in the case in which Q is in the actual world the underlying physical property of objects which are red.

"But there is no such meta-level" (*Ruling Passions*, p. 311). We need only versions of the Fixedly-Actually operators to express what Blackburn thinks cannot be expressed.

This discussion should also make clear the strict limits of the earlier concession to mind-dependent theorists that allowed them to regard moral principles as judgementally valid. That concession can be granted only on the understanding that the first parameter, the attitudes that according to them are the source of moral truth, is held fixed.

We might pick out a moral principle not by its content, but by some definite description that relates it to those who accept it. The mind-dependent theorist does have access to some principle such as the following: in any world, a basic principle that is morally endorsed in that world will be one that holds in that world. This is identifying a moral principle by description, rather than by its content. Now a given matrix, of the sort I have introduced, corresponds to a statement identified by its content, by a that-clause. So the principle to which I have just agreed the mind-dependent theorist does have access is not a principle that ensures that in a *given* matrix, all the entries along the diagonal are "True". Rather, what it ensures is something concerning a set of many different matrices. It ensures that, for a given world w, if P is a statement endorsed by the basic morality-generating attitudes in w, then the entry in the matrix in the column labelled "Att_w" for the row for w will be "True". This does not ensure what is required by the status of a given proposition as contentually a priori, namely, possession of the entry "True" along the diagonal of a single given matrix. Rather, it gives only something weaker. It gives a diagonal of "True" entries in three dimensions, if you will, across a series of different two-dimensional matrices.

My position, in contrast to all mind-dependent views of moral principles, is that there is no sense in which moral principles fail to be contentually a priori. I hold this to be an epistemic and metaphysical truth. It is not itself a moral truth. The trouble for mind-dependent theorists is caused by variation with respect to the first parameter in P(w1, w2).[23] If any form of mind-dependent theory of moral judgement is correct, that parameter must be articulable, at least at the level of philosophical reflection. My own view is that a proper appreciation of the contentually a priori status of moral principles ought to lead us to believe that any such parameter or argument-place is otiose. The moderate rationalist about morality who is also tempted to some form of subjectivism about colour will say that while basic moral principles are contentually a priori, so that—if the relativization is insisted upon—(DM) is true, in the case of colour, the characteristic consequence of one form of subjectivism holds, in that (DC) is false.

It is not in fact my view that our basic moral prima facie principles could intelligibly have been utterly different, in ways which have no connection with rationales

[23] The problem is specific to this feature. There is of course no general incompatibility between a domain of truths—such as the truths about which material objects have which colours—being mind-dependent and the existence of contentually a priori truths about the properties attributed in that domain. Principles of colour incompatibility are contentually a priori (see the explanation suggested for their being so in "Explaining the A Priori"). This is not merely consistent with truths about colour being significantly mind-dependent. In my judgement, the explanation of their status as contentually a priori has to draw upon the special relation of colour concepts to colour experience.

for the principles we in fact accept. That possibility was being entertained in the preceding part of the argument only for ad hominem purposes. My claim is that the mind-dependent theorists do not have the resources to rule out such variation, and so cannot explain why basic moral principles are contentually a priori.

In this discussion, I have focused on the formulation used by Blackburn; but in fact the points I have been making seem to apply to any subjectivist or mind-dependent theory that tries to avoid the problems by using an "Actually" operator. Subjectivist and mind-dependent theorists are naturally tempted to appeal to our actual subjective states, or judgements, and to say that modal propositions about the moral should be evaluated always with reference to those actual states or judgements.[24] Contrary to the views of many writers in this area, I myself think that a proper deployment of the formal modal apparatus all things considered tells against mind-dependent approaches to morality, in a way in which it does not tell against mind-dependent approaches to statements about colours.[25]

There are links and affinities between the Sharpened Thesis and G. E. Moore's justly famous paper "The Conception of Intrinsic Value".[26] Moore was very opposed to the idea that the goodness of something could be a matter of its extrinsic, rather than its intrinsic, properties. He was equally opposed, whether the extrinsic properties in question were conceived of as mind-dependent or were conceived of as mind-independent. But in the special case in which mind-dependent qualities were offered as an analysis of goodness, he wrote of "the fact that, on any 'subjective' interpretation, the very same kind of thing which, under some circumstances, is better than another, would, under others, be worse—which constitutes, so far as I can see, the fundamental objection to all 'subjective' interpretations" (p. 283). This formulation is of course ineffective against the subjectivist who uses "Actually" operators; but it would be a very superficial understanding of Moore which took this as a reason for saying that his "fundamental objection" fails. On Moore's view, as he stated it in italics, "To say that a kind of value is 'intrinsic' means merely that the question whether a thing possesses it, and in what degree it possesses it, depends solely on the intrinsic nature of the thing in question" (p. 286). There is clearly still dependence of value on a thinker's mental states if the subjectivist formulates his theory using "Actually" operators. The right way to demonstrate this dependence is not to appeal to

[24] For such use of an "Actually" operator in defending a subjectivist theory, see D. Wiggins, "A Sensible Subjectivism", in *Needs, Values and Truth: Essays in the Philosophy of Value* (Blackwell: Oxford, 1987), p. 206. Wright in *Truth and Objectivity* describes the use of an "Actually" operator as "an attractive strategy" (p. 114), but, rightly in my judgement, goes on to caution that "no proposition whose necessity is owing entirely to actualisations can be known *a priori*" (p. 116).

[25] Wiggins cites Davies and Humberstone, pp. 22–5, in support of his use of an "Actually" operator to meet the objections to subjectivism. D. Lewis gives a very clear acknowledgement of the problem for a subjectivist theory: "The trick of rigidifying seems more to hinder the expression of our worry than to make it go away. It can still be expressed . . .": "Dispositional Theories of Value", op. cit., at p. 132 ff.

[26] In his *Philosophical Papers* (London: Routledge & Kegan Paul, 1922), pp. 253–75, but now most accessible in the Revised Edition of *Principia Ethica*, ed. T. Baldwin (Cambridge: Cambridge University Press, 1993), pp. 280–98. Page references in the main text above are to this more recent volume.

metaphysical possibilities, but to the failure of this subjectivist's conditions to hold Fixedly Actually. There is still dependence of the sort to which Moore objects if the dependence is on thinkers' actual attitudes or other subjective states.

With this understanding of Moore's intentions, there is an intuitive argument from Moore's Thesis that moral values are intrinsic to the Sharpened Thesis.[27] We can argue by contraposition. If the Sharpened Thesis were false, then there would be moral principles, and so statements of value, that could be known only by empirical investigation of the actual world. It is hard to see how those values, thus knowable only empirically, would be intrinsic in the sense that mattered to Moore.[28]

If we step back to reflect on the argument I have given so far, it is apparent that it does not depend on features that are unique to morality. The argument I have offered so far can be developed in corresponding form to reject any mind-dependent treatment of any domain in which there are principles that are contentually a priori. The argument could be applied against mind-dependent treatments of metaphysical necessity, for instance (if further arguments against such treatments were thought to be needed). All the arguments against mind-dependent treatments in the moral case would carry through *pari passu* for the modal case. One could even imagine a G. E. Moore-like philosopher writing a paper called "The Intrinsic Conception of Necessity", in which the author insists that whether a proposition is necessary depends only upon the nature of its various constituents, and does not depend on any thing external to those constituents, whether it be mental or non-mental.

3. Explaining the A Priori Status of Morality: A Schema

The claim that basic moral principles are contentually a priori does not by itself imply the view that they can be derived from the law of non-contradiction. The laws of modal logic, and other basic principles of metaphysical necessity, are also a priori. But they are not literally derivable from the law of non-contradiction alone. Otherwise modal logic would be a part of first-order logic, which it is not. Kant himself of course believed in a connection between what you can will without contradiction and the correctness of a principle. But his *Groundwork* also contains another idea, a more general idea, which does not in its basic formulation mention non-contradiction. This more general idea contains the seeds of an explanation of the a priori status of moral principles. Kant writes:

> the ground of obligation here must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in concepts of pure reason.[29]

[27] Moore writes about values more generally, including aesthetic value. Obviously the Sharpened Thesis, restricted as it is to moral values, could have consequences at most for moral values.

[28] The converse implication holds only under the additional supposition that any extrinsic property is not knowable a priori. This supposition would not be true for arithmetic. It is not an intrinsic property of the number 4, but it is knowable contentually a priori, that it is the minimum number needed to colour an arbitrary map on the plane without adjacent regions having the same colour. The notions in play at this stage of the discussion would thus need some refinement for this converse implication to be established. I conjecture that such refinement is possible.

[29] *Groundwork* 4: 389, p. 45 in the Cambridge edition, op. cit.

This claim of Kant's is a consequence of the highly plausible principle that ways of coming to know a given proposition that are a priori ways have their source in the nature of one or more concepts in the given proposition. This principle is part of the moderate rationalism I mentioned at the start of this paper. If moral principles are a priori, and a priori ways of coming to know a proposition trace back to the nature of the concepts it contains, it follows that some ways of coming to know a moral principle have to do with the nature of moral concepts. Our task is to say how this is so.

I have already mentioned the modal case twice, and it will continue to help us to consider the partial parallel between modal and moral concepts. As I said, modal truth seems to be fundamentally contentually a priori, like basic moral principles. Elsewhere, I argued that our understanding of modal truth is best explained by our having an implicit conception whose content is given by a set of principles that collectively determine which world-descriptions represent genuine possibilities.[30] Those principles I called the "Principles of Possibility". The Principles of Possibility, whose details do not matter for present purposes, include principles entailing that genuine possibilities respect what is constitutive of the identity of the concepts, object, properties and relations they concern. What matters in considering a partial parallel with the moral case is the model of understanding, epistemology and metaphysics instantiated by this principle-based approach. Under the principle-based approach, to understand modal operators is to evaluate modal claims as true or false in accordance with these principles. The principles are at most tacitly known to an ordinary thinker when she evaluates modal claims. It takes philosophical thought to work out what those principles are.

The principle-based approach to modality has two features that we equally need to provide for in the moral case.

First, it gives an account of how a way of coming to know, even one employed by a non-philosophical thinker, can be a way that ensures that what is known is true in the actual world, whichever is the actual world. In evaluating modal claims, the thinker draws on the content of tacit knowledge of the Principles of Possibility. These Principles state what it is, constitutively, for a description to represent a genuine possibility. The Principles are themselves true in the actual world, whichever is the actual world—they hold Fixedly Actually. Standard logical inferences will preserve Fixedly-Actual truth. Truths about what is constitutive of particular concepts, objects and properties are equally plausibly truths that hold Fixedly Actually. If our thinker draws only on information which holds Fixedly Actually, by rules which preserve that property, when she evaluates modal truths, then the modal truth she comes to know thereby will hold Fixedly Actually. This is a way of coming to know a modal truth that ensures that what is known will hold in the actual world, whichever is the actual world.

This general method of evaluating modal claims is not infallible—such general methods never are. A thinker may make mistakes about what is constitutive of the identity of a concept, object, property or relation; she may also make inferential mistakes. But when there are no such mistakes, the way in which a modal belief is

[30] *Being Known*, chapter 4.

reached can be one ensuring that its content is true in the actual world, whichever is the actual world. Since the existence of such ways of coming to know contentually a priori modal propositions relies on an account of understanding modal notions, and does not involve causal interaction with a modal realm, the principle-based account is a species of moderate rationalism for the modal case.

The other feature of the principle-based approach to modality that we also need to provide for in the moral case is its provision of a straightforward means for integrating the modal epistemology and modal metaphysics that steers between the extremes of mind-dependence on the one hand, and an epistemology that requires causal contact with a modal realm on the other. If the Principles of Possibility state what it is for something to be a genuine possibility, and those Principles are properly applied in reaching modal beliefs, we already have an explanation of how modal knowledge is possible. Such a middle course, avoiding both mind-dependence and interactionism, is just what we need in the case of morality too.

The moral analogue of the principle-based treatment of modality is a treatment under which to possess moral concepts involves having an implicit conception whose content is operative when one assesses moral propositions. Full grasp of a given moral concept, if such a thing is ever possible, would involve possession of an implicit conception whose content formulates what it is, constitutively, for something to fall under that moral concept. The general idea of a principle-based treatment is in itself neutral on what the content of the implicit conceptions are. Many different first-order moral views could avail themselves of a principle-based treatment in attempting to address epistemological and metaphysical issues about the status of morality. So equally could many different philosophical views about what unifies the principles that form the content of the implicit conceptions. I will not be taking on the task of addressing particular first-order moral views here, nor the question of what unifies them. My aim is rather to consider what resources a principle-based treatment makes available to a variety of conceptions when they turn to address epistemological and metaphysical issues.

The implicit conceptions possessed by a moral thinker will be complex and structured. They will concern values, ideals, their relative importance, and something about their underlying sources. Even from a description as brief as that, there are two apparent differences from the modal case. One of the most important differences is the need for some kind of "prima facie" or "pro tanto" operator in the moral case, which, in my view, has no analogue in the modal case. It is plausible that one will need to employ, in any principle-based account of moral truth and moral epistemology, principles of the form "Prima facie, given that an action is F, it is good in such-and-such a respect". The same applies to evaluations of states of affairs. The presence of a prima facie operator has many repercussions, including some for the issue of determinacy. There is nothing in such structures to rule out the possibility that some type of action may be prima facie good in certain respects, prima facie bad in others, and there be nothing further in the principles to settle outright whether it is good or bad.

A second difference from the modal case concerns completeness. There is some plausibility that we can give a very general characterization of what is required for

a description to represent a genuine possibility. It is arguable that if a description respects what is constitutive of concepts, objects and properties, it represents a genuine possibility. Though we are certainly ignorant, for many concepts and objects, of what it is that is constitutive of them, such ignorance concerns whether the conditions for certain possibilities are met, and is not about what it is for something to be possible. It is not apparent that anything analogous has to hold in the case of moral thought. Even our implicit conceptions may be incomplete, may need further articulation from reflection on examples and other principles.

A thinker may have an implicit conception with a correct content involving a given concept, but nevertheless make mistakes when asked to formulate general propositions involving that concept. This is a familiar phenomenon of implicit conceptions in other domains, evidenced by the frequent inability of thinkers to, say, define "chair" correctly, or to state explicitly the rules of grammar they are following. Indeed, even the simple example I have been using needs qualification. The infliction of avoidable pain is not prima facie wrong in the case in which the pain still exists, but is not experienced as hurting, as is the case for one who has taken morphine. The infliction of pain is prima facie wrong only when it is a form of suffering, and is wrong for the same reason as it is wrong to cause, say, avoidable depression or severe anxiety in a person. Reflection on the ways in which we can correct our initial impressions of wrongness or rightness will make the principle-based theorist say that not all cases are like those which Prichard described as immediate apprehension, in which "insight into the nature of the subject directly leads us to recognize its possession of the predicate".[31]

A thinker who judges that some type of action is wrong may be more or less articulate in his ability to say why it is so. At the least articulate level, the thinker may just make some clear intuitive judgement that it is wrong, without being at all confident in any particular explanation of why it is wrong. At one step up from this, the thinker may be able to give a ground: "because it would be a betrayal", "because it hurts him and the hurting is avoidable". At another step up, the thinker may be able to say why these are grounds. Higher levels of justification involve abductions from a priori examples and other apparently a priori principles—at this level of description, the methodology is the same as that found in other domains in which truth is fundamentally a priori. The possibilities of error are the same as in other a priori domains.

A principle-based approach can share each of the features that made the parallel to the modal case tempting. Even if a thinker's implicit conception of some moral property is incomplete, the content of that conception can still be a correct partial statement of what it is, constitutively, for something to fall under that concept. They will, for instance, be a correct partial statement of what determines the semantic value of a concept like *is prima facie wrong*. When they are so, and when the information is properly drawn upon in the evaluation of contents containing that concept, the contents thus reached will be true. And as in the modal case, since the rule determining

[31] H. A. Prichard, in his essay "Does Moral Philosophy Rest on a Mistake?", repr. in *Moral Obligation and Duty and Interest: Essays and Lectures* ed. J. O. Urmson (Oxford: Oxford University Press, 1968), at p. 8.

the semantic value of a concept applies whichever world is the actual world, propositions thus reached will be reached in a way that guarantees that what is known in that way will also hold in the actual world, whichever is the actual world.

The fact that implicit conceptions are involved in the evaluation of moral propositions does not by itself suffice to account for the contentually a priori status of basic moral principles. There is no contradiction in the idea of an implicit conception having an a posteriori content. In fact, an implicit conception with the content that the word 'chair' in one's own language applies to things having certain properties is an implicit conception with an empirical content. What matters for a priori status is rather that the given way of coming to know is guaranteed to be correct by the way in which the semantic values of the relevant concepts are fixed. Implicit conceptions whose contents either consist of principles that state what it is for something to be wrong, for instance, or consist of consequences thereof, meet this further condition. Without this further condition, we would not have an explanation of the contentually a priori status of basic moral principles. The same applies to the modal case.

This integration of the metaphysics of the moral—what it is to fall under certain normative concepts—with an epistemology also steers the same middle course as the principle-based account of modality. It involves neither a mind-dependent account of moral truth, nor a causal epistemology for the contentually a priori principles.[32]

The point that fundamental principles help to determine the semantic value of concepts like *is prima facie wrong* is important in separating any principle-based conception from mind-dependent treatments of moral thought. Mind-dependent theorists can fairly insist that on their views, a certain set of moral principles is correct, and can equally insist that some principles are more fundamental than others. That does not imply that mind-dependent theorists can simply take over the apparatus of the principle-based view. The objection remains outstanding against the mind-dependent theories that they cannot explain the contentually a priori status of basic moral principles. To try to meet the objection by saying that the principles themselves determine the semantic value of moral concepts, regardless of what attitudes minds take to them, would be to abandon any claim of mind-dependence. A principle-based conception is a very different animal from any mind-dependent view.

4. The Subjectivist Fallacy

The Subjectivist Fallacy is the fallacy of moving from a premiss stating that certain mental states are sufficient, or stating that certain mental states are necessary, for a given content to be true, to the conclusion that the truth of the content consists, at least in part, in something subjective or mental. I say that this is a fallacy even in the case in which the premiss stating that certain mental states are sufficient, or are necessary, holds true a priori. To say that it is a fallacy is not of course to say that the conclusion is not true: only that it cannot be supported just from these premisses.

[32] It does not follow that moral properties may not be involved in other causal explanations, not having to do with knowledge of a priori principles.

The Subjectivist Fallacy is a fallacy because it may be possible to explain why the mental states are necessary or sufficient for the truth of the target content by exhibiting this necessity or sufficiency as a consequence of a more fundamental account of what is involved in the truth of the target content, a more fundamental account that does not mention mental states at all. The fact that there is in a certain sense no gap between certain mental conditions obtaining and the holding of the target content may have a non-subjectivist explanation.

Here is an example of the Subjectivist Fallacy, an example which would be recognized as such on all but the most extreme views of the nature of meaning and rule-following. The case involves a hypothetical position on the understanding of arithmetical relations. We can imagine a theorist who starts from this true premiss:

> Within the accessible numbers, it is sufficient for $n + m$ to equal k that a thinker who reaches his judgement about what $n + m$ equals in accordance with certain recursive procedures will judge that $n + m = k$.

From this truth, our imagined theorist moves to the conclusion

> Equations involving addition have partially mind-dependent truth-conditions concerning what a certain kind of thinker would judge.

Almost everyone will agree that this hypothetical theorist's mistake lies in not realizing the judgements of his hypothetical calculating subject are correct only because they respect the recursive equations for addition. The fact that there is (and is a priori) a necessary and sufficient condition, framed in terms of the judgements of a hypothetical thinker, for the holding of the addition relation on the accessible numbers is just a by-product of something more fundamental. This more fundamental condition is the non-psychological truth-condition for equations involving addition determined by the recursive characterization of the addition relation.

How does this bear on constructivism in ethics? Constructivists need not be mind-dependent theorists. Constructivists too can agree that the displayed transition about addition moves from a true premiss to a false conclusion, provided their constructivism is of a non-psychological variety. Their constructivism will be of a non-psychological variety only if the "can" in the phrase "can be constructed" which features in a statement of constructivism is not explained in psychological terms. It is also a necessary condition of the constructivism being non-psychological that the particular rules or recursions it mentions are not mentioned there by virtue of their meeting some mind-dependent condition. Some versions of constructivism meet these conditions. Hence constructivists need not be mind-dependent theorists.

A second example of the Subjectivist Fallacy moves from truths about concept-possession to a general subjectivism about truth. The premiss of this second illustration of the Subjectivist Fallacy is available to anyone who accepts two points about concept-possession. The first point is that in the possession condition for a concept, reference is made to what the thinker must be willing to judge in certain circumstances. If such reference is thought to be required only in certain cases, then the premiss of this illustration will correspondingly be available only in that restricted class of cases. The second point needed for the availability of the example concerns

the theory of concepts and the theory of the way in which their reference is determined. These two theories must, one way or another, jointly have the consequence that the judgements a thinker is required to make in given circumstances, if he is to be credited with possession of the concept, are ones which are true in those same circumstances. (On some theories, this second point is secured by the account of how semantic values are assigned to concepts.) In any case, if these two points are accepted, perhaps just for a restricted range of concepts φ, then a premiss of the following form will be correct:

If the thinker judges in the given circumstances that something is φ, then it is φ.

Does it follow that something's being φ, at least in the specified circumstances, is a mind-dependent, subjective matter? It certainly does not. The possession condition framed in terms of willingness to judge may determine a property an object has to have if it is to fall under the concept φ, a property that may, for all that has been said so far, be wholly mind-independent. In the case in which the concept is one of a logical constant, the property determined may be (or determine) a certain-truth function. In any case in which such a property is determined, the truth of the premiss will be explicable from a non-subjective account of the truth of contents of the form "a is φ". The same point applies when the property determined is one that would classically be recognized as a primary quality.

The more general fallacy described in the second illustration actually has the first illustration, about elementary addition, as a special case. It is the special case in which the concept is that of addition, and the possession condition for the concept requires computational practices which obviously respect the recursion which defines addition.

One element in Wittgenstein's rule-following considerations—perhaps not the only element, but an important and extensive element—is that justification comes to an end, and that in an account of understanding at a certain point we have to speak of an ability to go on in the right way. This is captured in possession-conditions which speak of what the thinker finds primitively compelling, without proceeding through inferential justifications for applying the concept in question. I have argued that we can recognize the subjectivist fallacy as a fallacy whilst employing such possession-conditions for concepts. If that is correct, it follows that there is nothing in this element of the rule-following considerations to rule out the more rationalist conception of moral thought that I was suggesting earlier.[33] Only on much more radical views of rule-following, for instance the view that the correctness of a judgement is not fundamentally the result of two components, the way the world is and the nature of the concepts employed, would such general conclusions of mind-dependence be acceptable.

The crucial step in the subjectivist fallacy as I have described it is acceptance of an incorrect criterion for the mind-dependence of a given property. Hence it is possible

[33] For an expression of sympathy with certain Wittgensteinian views in support of a treatment of moral thought that is very distant from the present rationalist view, see B. Williams, "Philosophy as a Humanistic Discipline", *Philosophy* 75 (2000), 477–96.

to make what seems to me the same mistake as is made in the subjectivist fallacy without actually being a subjectivist. Even a theorist who rejects subjectivism about a given domain may still be using a questionable account of mind-dependence of a given property. The theorist may even be relying on that account in his rejection of subjectivism. The writings of Crispin Wright and Mark Johnston contain examples of criteria for mind-dependence that seem to me open to question in this way. In his well-known discussion of the Euthyphro Contrast, Wright introduces the notion of a "provisional equation", which is something having the form of a conditional whose consequent is itself a biconditional, i.e. the form $A \supset (B \equiv C)$. A provisional equation is something of the form "If CS, then (it would be the case that *p* if and only if S would judge that *p*".[34] A substantial provisional equation, says Wright, has an antecedent CS in which "a concrete conception is conveyed of what it actually does take" for the subject to be operating under conditions in which her opinion is true.[35] Wright endorses this conditional: "if a discourse sustains substantially formulated true provisional equations which can be known a priori to be true, then that makes the beginnings of a case for regarding the discourse as dealing in states of affairs whose details are conceptually dependent upon our best opinions".[36] Similarly Mark Johnston, in addressing the question "How then are we to demarcate the response-dependent concepts?" offers the answer that if a concept *C* is one interdependent with, or dependent upon the responses of subjects, "then something of the following form will hold *a priori*

> *x* is *C* iff in K, Ss are disposed to produce *x*-directed response *R* (or *x* is such as to produce *R* in *Ss* under conditions *K*)."[37]

This biconditional will be fulfilled in our first, arithmetical, example, when we take the concept *C* to be the property of (say) being the sum of 7 and 5, the condition K to be the condition of exercising properly-functioning memory and perceptual systems, and the response to be that of making a certain judgement expressing the outcome of the subject's computation in accordance with certain rules. A corresponding point could be made about Wright's criterion. In both Johnston's and Wright's proposals, the test proposed for mind-dependence is too easily met by propositions whose truth is not mind-dependent. Nothing can be validly concluded from the existence of such a priori conditionals or biconditionals in a given domain about the mind-dependence of that domain.

There is nothing inimical, in these illustrations and arguments, to the idea that some contents do have mind-dependent truth conditions. They do when their truth-conditions concern a property whose nature—what it is, constitutively, to have to property—is to be explained in terms of properties of the mind. The burden of the preceding remarks is that this constitutive condition cannot be reduced to something involving a priori equivalence with conditions concerning certain mental states.

[34] *Truth and Objectivity*, p. 119. I have altered only Wright's notation for propositions.
[35] Ibid., p. 112. [36] Ibid., pp. 119–20.
[37] "Dispositional Theories of Value", *Proc. Arist. Soc. Supp. Vol.* LXIII (1989), 139–74, p. 145.

The Subjectivist Fallacy is an instance of a more general fallacy concerning the nature of properties. The more general fallacy is that of moving from the a priori truth of a biconditional of the form

F(x) iff A(x)

to the conclusion that being A is what *makes* something F. I call this "the Biconditional Fallacy". Just as in the subjectivist case, it is a fallacy because the correct account of what makes something F may have a consequence that it is a priori that something is F iff it is A; but the constitutive account may not mention properties or notions of the sort mentioned in the condition A(x). One of the tasks facing those who want to develop Discourse Ethics, for example, is to show that it can be done without committing this fallacy. Habermas formulates the central claim of discourse ethics as follows: "Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse".[38] Let us suppose that the theorist of discourse ethics makes a good case that this condition is a priori true. Nothing would follow about what makes something a valid norm. The approval, in the appropriate practical discourse, of those affected might be a consequence of more fundamental principles about norms that have this approval as a consequence. Consistently with the principle Habermas formulated being a priori, practical discourse might not be mentioned in an account of what is fundamentally constitutive of the notion of a valid norm. It is that further claim about fundamental constitution that Discourse Ethics would have to establish if it is to speak to the nature of morality.

I conclude this section with a more general reflection on the theoretical options available to us. When one reads the literature on judgement-dependent and other mind-dependent approaches to ethics and other subject-matters, the impression is often conveyed that, when we do have an a priori biconditional linking some property with thinkers' mental states, there are only two options. Either we read the psychological material of the right-hand side as providing what is constitutive of the left-hand side's holding: or else we must accept some form of "detectivism", with the overtones of "detection" involving a causal epistemology for the states of affairs detected.[39] To think that these are the only two possibilities is to overlook broadly rationalist approaches that are neither mind-dependent nor committed to the possibility of causal interaction. It is as if the only two possibilities in the philosophy of arithmetic, or the philosophy of modality, were either subjectivism or a commitment to causal interaction. I suggest that a good rationalist treatment of mathematics, modality and morality involves neither of those two positions, but genuinely presents a third way.

[38] *Moral Consciousness and Communicative Action*, tr. C. Lenhardt and S. Weber Nicholsen (Cambridge, Mass.: MIT Press, 1990), p. 93.

[39] See *Truth and Objectivity*, on the Euthyphro contrast, Appendix to chapter 3.

10

Indexical Concepts and Compositionality

François Recanati

Indexical expressions are characterized by their two-dimensional semantics. They have a 'content', that is, an intension in the traditional sense: something that determines the extension of the expression, given a situation of evaluation. But they also have a *primary* intension, or 'character': something that determines the expression's content, given a situation of utterance.

For simple indexicals, as opposed to complex indexical phrases such as 'my sister', the content is a constant function. Simple indexicals are conventionally associated with a rule which, in the situation of utterance, fixes the extension directly, in such a way that it does not vary with the situation of evaluation. (As Kaplan says, the reference is fixed in context 'before the encounter with the circumstance of evaluation'.) The rule associated with 'I' is the rule that a token of 'I' refers to the person who utters that token. The rule associated with 'here' is the rule that a token of 'here' refers to the place where the token is uttered. In all cases the reference is the entity which stands in the right contextual relation to the occurrence of the expression. What is conventionally encoded in the expression-type, independent of context, is the nature of the contextual relation in question; but the entity which contextually stands in that relation to the token is what it (the token) contributes to the possible-worlds truth-conditions of the utterance. In the two-dimensional framework, simple indexicals can be treated as rigid designators—that is, expressions which refer to the same entity in all possible worlds of evaluation—despite the fact that their reference depends upon, and varies with, the situation of utterance.

In what sense can we talk of mental indexicality? Linguistic conventions have no role to play here. Still, I hold that some (simple) concepts are indexical in the sense that they too are associated with a rule which contextually determines the reference. The reference of such a concept is the entity which stands in the appropriate contextual relation to the thinker in whose thought the concept occurs. That entity is what the concept contributes to the truth-conditional content of the thought, while the nature of the contextual relation in question determines the *type* of the concept (its cognitive role). So indexical concepts are susceptible of the same sort of two-dimensional analysis as indexical expressions.

In the first part of this paper I will sketch a theory of indexical concepts within a broadly epistemic framework.[1] In the second part I will discuss and dismiss an argument due to Jerry Fodor, to the effect that any epistemic approach to concept individuation (including the theory of indexical concepts I will sketch) is doomed to failure.

1. Indexical Concepts: An Overview

1.1 What indexical concepts are, and what they are for

Following Strawson, Perry and others, we can think of concepts—in many cases at least—as mental files in which we store information concerning the extension of the concept. Thus my concept LION is a file in which I store what I know or believe regarding lions, and my concept GEORGE W. BUSH a file in which I store what I know or believe regarding Bush. Indexical concepts can be construed as special files whose very existence is contingent upon the existence of certain contextual relations to entities in the environment. The file exists only as long as the subject is in the right relation to some entity; a relation which makes it possible for him or her to gain perceptual information concerning that entity. Thus in virtue of being a certain person, I am in a position to gain information concerning that person through, for example, proprioception. The mental file SELF serves as repository for information gained in this way. The concept HERE which occurs in my current thoughts concerning this place is a temporary mental file dependent upon my present relation to the place in question. I occupy this place, and this enables me to gain information concerning it simply by opening my eyes and my ears. The perceptual information thus gained goes into the temporary file.

When the contextual relation on which the information link depends no longer exists, the file/concept is suppressed. When I leave this room, I can no longer think of this room as HERE; I have to think of it under a different concept. I can still think HERE-thoughts, but the HERE-concepts occurring in those thoughts will be concepts of different places, hence different concepts (though concepts of the same type as my present HERE-concept).

I assume that demonstrative concepts, such as the concepts THAT MAN or THAT THING, are a subclass of indexical concepts. They are based on certain contextual relations to objects, in virtue of which we can not only perceive them but also focus our attention on them in a discriminating manner. When we are no longer in a position to perceive the object or to focus our attention on it, we can no longer think of it under the demonstrative concept which depends upon the existence of a suitable demonstrative relation.

1.2 Cognitive dynamics

Indexical concepts, as I said, are mental files in which we store information gained via the contextual relations on which the concept is based. But what happens when the

[1] See Recanati (1993), chapters 6 and 7, for an elaboration. The foundations of the theory can be found in the work of Gareth Evans and John Perry (see Evans 1982, Perry 1993).

relation is broken and the temporary file based on it disappears? What happens to the information stored in the file?

A similar question arises with respect to indexical expressions. When the context changes, we cannot express the same content unless we adjust the indexicals to the new context. As Frege said,

> If someone wants to say today what he expressed yesterday using the word 'today', he will replace this word with 'yesterday'. Although the thought is the same its verbal expression must be different in order that the change of sense which would otherwise be effected by the differing times of utterance may be cancelled out. (G. Frege, 'Thought', in Beaney 1997: 132)

Similarly, an adjustment of indexical concepts must take place if the context changes. As I pointed out earlier, I can no longer think of a place as HERE if I no longer occupy that place. And I cannot think demonstratively of an object which I can no longer perceive. In both cases, however, another indexical concept is readily available. In the demonstrative case, the demonstrative relation to the object no longer holds, but another relation holds, in virtue of which I remember the object. On that relation another indexical concept is based, distinct from but closely related to the original demonstrative concept. Following Evans (1982), let us call the new concept a 'past-oriented demonstrative', or 'past demonstrative' for short. Just as demonstrative concepts (or 'present demonstratives') are based on demonstrative relations in virtue of which one can perceive the object, past demonstratives are based on certain relations in virtue of which one can remember the object.

To sum up, when a demonstrative concept comes out of existence because the demonstrative relation on which it is based no longer holds, a past-demonstrative concept systematically comes into existence because the perceptual episode has impressed our memory. Through our memories of the object, we can focus our attention on it even after the perceptual encounter has ended. We can therefore say that the present demonstrative THAT MAN [WHOM I SEE] is *converted into* a past demonstrative THAT MAN [WHOM I SAW].

Not only can an indexical concept be converted into another type of indexical concept, as in this particular case; two distinct indexical concepts can also be linked together. This is what happens when, for example, the subject recognizes a certain object which he perceives as being a certain object which he has perceived before and still remembers. In recognition, a demonstrative concept and a past demonstrative are linked together. This linking gives rise to a third type of concept based on a more complex relation which I call 'familiarity'. An object is familiar to the subject whenever multiple exposure to that object has created and maintained in the subject a disposition to recognize that object.

1.3 Recognitional concepts

Some concepts are based on the familiarity relation; I call them recognitional concepts (with apologies to those who use that phrase in a broader sense). A striking feature which distinguishes recognitional concepts from demonstrative concepts is that they are stable: they depend upon the continued existence of the subject's disposition to recognize the object, which disposition transcends particular encounters

with the object. Despite this stability, recognitional concepts are indexical, I claim. First, they depend for their very existence upon the existence of a contextual relation to the object, namely the relation of familiarity. Second, the reference of a recognitional concept depends upon the context: it is that object (if any) multiple exposure to which has created and maintained in the subject the recognitional disposition which underlies the concept. Which object that is depends upon the context. In a different environment, the very same recognitional device in place in the subject would have had the function of detecting another object than what it actually has the function of detecting in the actual environment.

Natural-kind concepts are themselves recognitional concepts, distinguished from the above by the fact that their content is arguably general rather than singular. We use the superficial or 'stereotypical' properties of water to detect water in the environment. What we detect is that substance (H_2O) multiple exposure to which has created and maintained in us the disposition to recognize it. But in a different environment a different substance would possibly play the same role: it would have the same superficial characteristics and multiple exposure to it would have created and maintained in us the same disposition to recognize it via those characteristics. In such a context we would have a concept very similar to our WATER-concept and internally indistinguishable from it, but it would not be a concept of water. It would be a concept of twater or XYZ (however we call the substance which plays the role of water on Twin-Earth). The reference of our WATER-concept therefore depends upon the context, even if the context at issue is much broader than the context relevant to determining the reference of HERE. In this way Putnam's claim that natural-kind concepts are indexical can be justified.

1.4 Deferential concepts

Like demonstrative concepts, recognitional concepts presuppose some form of acquaintance with the reference, hence the extension of the notion of indexical concept which I have just suggested may seem natural. But what about cases in which the subject is not acquainted with the reference but has merely second-hand knowledge of it? I have argued that, in such cases, the subject possesses a *deferential concept*, and that deferential concepts themselves are indexical (Recanati 1997, 2000a, 2000b, 2001). While the indexical concepts talked about so far serve as repository for information gained in perception through various relations of acquaintance with the reference, deferential concepts serve as repository for information gained in communication through *linguistic* relations to the reference.

My hypothesis is that there is, in the mental repertoire, a 'deferential operator' which enables us to construct deferential concepts with a two-dimensional semantics analogous to that of the indexical concepts we have dealt with so far. The deferential operator $R_x(\,)$ applies to (the mental representation of) a public symbol σ and yields a mental representation $R_x(\sigma)$—a deferential concept—which has both a character and a content. The character of $R_x(\sigma)$ is basically the following rule for the determination of content:

(DO) The content of $R_x(\sigma)$ = the content of σ, when used by x

That character is a function from contexts in which there is a user *x* of σ (implicitly referred to by the speaker/thinker) to the contents which σ takes when used by *x* (given the character *x* attaches to σ). What is special with the deferential concept $R_x(\sigma)$ is that *its* content is determined 'deferentially', via the content σ would take if used by *x*.

There is something clearly metalinguistic about deferential concepts. They involve tacit reference to the use of σ by *x*. But that metalinguistic aspect is located in the character of the deferential concept and does not affect its content. In virtue of (DO) the content of the concept $R_x(\sigma)$ as used by John is the same as (and is no more metalinguistic than) the content of the symbol σ when used by *x*.

Deferential concepts allow us to think and talk about matters we have no first-hand knowledge of. Even if I do not know what quarks are, I can (in speech, but also in thought) use the word 'quark' deferentially and thereby refer to quarks. This, of course, is possible only if there are competent users of the word 'quark' around for me to defer to. Thanks to deferential concepts, we are freed from the limited context of our own experience; but the content of our thought is still dependent upon the (linguistic) environment in which we live.

2. Compositionality and Epistemic Properties

2.1 Fodor's argument

I have claimed that certain epistemic relations to the referent are constitutive of indexical concepts, which are based upon those relations and exist only as long as they exist. But Fodor has repeatedly argued that *nothing epistemic can be essential to or constitutive of any concept*. This holds in virtue of a constraint which Fodor dubs the Compositionality Constraint (CC):

> (CC) Nothing can be essential to or constitutive of a concept unless it composes.

A property of a concept is said to compose just in case it satisfies the following condition: a concept has that property iff the concept's hosts (that is, the complex concepts of which it is a constituent) have it as well.

Insofar as the possession conditions for a concept are constitutive of that concept, (CC) entails that '*P* is a possession condition on a constituent concept iff it is a possession condition on that concept's hosts' (Fodor 2001a: 142). This biconditional is one of the many applications of the Compositionality Constraint. It is supported by the following consideration: If it is false, Fodor says, 'the following situation is possible: The possession conditions for RED are *ABC* and the possession conditions for RED APPLE are *ABEFG*. So denying [the Compositionality Constraint, as applied to possession conditions] leaves it open that one could have the concept RED APPLE and not have the concept RED' (Fodor 1998a: 37). But this is incompatible with the usual compositional account of productivity and systematicity. According to that account, RED APPLE is a complex concept *containing* RED *as a constituent*, and the semantic value (reference) of the complex concept is a function of the semantic values of its constituents. It follows that it should *not* be possible to have the concept RED APPLE without

having the concept RED. Fodor concludes that we need (CC) to explain the productivity and systematicity of concepts.

From (CC), it follows, according to Fodor, that epistemic properties cannot be essential to concepts, because epistemic properties precisely do not compose. Thus consider WATER. I have suggested that it is a recognitional concept, based upon a capacity to recognize water (in normal conditions). But that epistemic property supposedly characteristic of recognitional concept does not compose. Complex concepts such as that of WATER TANK are not themselves based upon a capacity to recognize water tanks in normal conditions. Or, if they are associated with such a capacity, that is accidental in the sense that the capacity in question—to recognize water tanks in normal conditions—does not itself depend upon the capacity to recognize water in normal conditions. Since epistemic properties do not compose, they are not essential to concepts and cannot be used to individuate them or to type them (as I have done in the first part of this paper). So the argument goes.

2.2 An inconsistent triad?

I grant Fodor that, to account for productivity and systematicity, we need the following assumptions:

Constituency: Concepts are used as constituents of more complex concepts.

Compositionality of reference: The reference of a complex concept is determined by the references of its constituents (and the way the constituents are put together).

I also accept Fodor's claim that the epistemic property characteristic of recognitional concepts—the fact that such a concept is based upon a disposition to recognize its instances in normal conditions—does not compose, and that the same thing holds of epistemic properties in general. In contrast to the concept's reference, which is compositionally determined by the references of its constituents, there is a sense in which the epistemic properties of a complex concept are not determined by those of their constituents.

What I question is the gist of Fodor's argument: the transition from the non-compositionality of epistemic properties to the impossibility of construing them as essential to concepts. Once we realize that epistemic properties do not compose, Fodor says, we can no longer take them to be essential to concepts without threatening the usual account of productivity and systematicity. That is what I deny. I think there is no inconsistency in holding simultaneously that

[1] Epistemic properties do not compose.

[2] The usual account of productivity/systematicity (that is, the account based upon the two assumptions listed above) is correct.

[3] Epistemic properties are constitutive of certain classes of concepts (for example, indexical concepts).

In other words, I hold that epistemic approaches to concept individuation are compatible with the usual account of productivity and systematicity even if we accept that epistemic properties do not compose. Hence what I will do, in the last section

of this paper, is scrutinize Fodor's argument to the effect that [1]–[3] form an inconsistent triad.

2.3 Simple inheritance versus compositional inheritance

What is incompatible with the usual account of productivity and systematicity is the claim that one could have the concept RED APPLE without having the concept RED.[2] Fodor thinks this claim follows from [1] and [3] in the above triad, but he is wrong. He would be right only if [1] entailed the *non-inheritance* of epistemic properties from constituent to host. But [1] only says that epistemic properties *do not compose*. This, I claim, is different from saying that they are not inherited, in the simplest possible sense of the term.

To show that the epistemic properties that are constitutive of constituent concepts are inherited by their hosts (even if they do not compose) is a trivial matter. If the complex concept RED APPLE (or WATER TANK) has the concept RED (or WATER) as a constituent, and the concept RED (/WATER) has, among its possession conditions, an epistemic capacity *S* (for example, the capacity to recognize red things, or water, in normal conditions), it *immediately* follows that one cannot have the concept RED APPLE without having the concept RED and therefore without having the epistemic capacity *S* (simple inheritance). What does *not* immediately follow is this: that one cannot have RED APPLE without having *an epistemic capacity S* which is to red apple what S is to red*, namely, the capacity to recognize red apples in normal conditions (*compositional* inheritance). In other words: The constitutive epistemic properties of constituent concepts are perforce inherited by their hosts, yet they do not compose in the sense in which standard semantic properties such as reference compose. The reference of the complex concept RED APPLE (or WATER TANK) is compositionally determined by the references of its constituents. That implies that the complex concept *has* a reference of its own, which is determined by the references of its constituents. But the complex concept RED APPLE can inherit the epistemic possession conditions of its constituents *without having an epistemic possession condition of its own* (let alone one determined by the possession conditions of its constituents): again, one can have the concept WATER TANK without having the capacity to recognize water tanks; or, if one has the capacity to recognize water tanks, it will not be determined by one's capacity to recognize water in the way in which the reference of WATER TANK is determined by (*inter alia*) the reference of WATER.

Compositionality turns out to be a much stronger form of inheritance than what I called 'simple inheritance'. But only the failure of simple inheritance would threaten the usual account of productivity and systematicity, by forcing us to acknowledge the possibility of having RED APPLE without having the concept RED. In the relevant passages where he presents his argument against epistemic approaches to concept individuation, Fodor systematically trades upon the ambiguity of 'inherit' between the two notions I have distinguished—simple inheritance and compositional inheritance. His argument is fallacious because it rests on that ambiguity. The fact that

[2] More specifically, that claim is incompatible with the assumption I dubbed 'Constituency'.

epistemic properties do not compose is the fact that the epistemic properties of the constituents are not *compositionally* inherited by the hosts. Still, the epistemic possession conditions for the constituents *are* inherited by the hosts (though not 'compositionally'), and that is sufficient to guarantee that one cannot have a complex concept without having its constituents.[3]

References

Beaney, M. (ed.) (1997). *The Frege Reader*. Oxford: Blackwell.

Chalmers, D. (1996). *The Conscious Mind*. Cambridge, Mass.: MIT Press.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.

Fodor, J. (1998a). 'There are no recognitional concepts—not even RED'. Chapter 4 of *In Critical Condition*, Cambridge, Mass: MIT Press/Bradford Book.

——(1998b). 'There are no recognitional concepts—not even RED. Part 2: The plot thickens'. Chapter 5 of *In Critical Condition*, Cambridge, Mass: MIT Press/Bradford Book.

——(1998c). *Concepts. Where Cognitive Science Went Wrong*. New York: Oxford University Press.

——(2000). Replies to Critics. *Mind and Language* 15: 350–74.

——(2001a). 'Doing without what's within. Fiona Cowie's *What's Within? Nativism Reconsidered*'. *Mind* 437: 99–148.

——(2001b). 'Language, thought and compositionality'. *Mind and Language* 16: 1–15.

Kaplan, D. (1989). 'Demonstratives'. In J. Almog, H. Wettstein, and J. Perry (eds.), *Themes from Kaplan*, New York: Oxford University Press, pp. 481–563.

Peacocke, C. (2000). 'Fodor on concepts: philosophical aspects'. *Mind and Language* 15: 327–40.

Perry, J. (1993). *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell.

——(1997). 'Can we believe what we do not understand?' *Mind and Language* 12: 84–100.

——(2000a). 'Response to Woodfield: deferential concepts'. *Mind and Language* 15: 452–64.

——(2000b). *Oratio Obliqua, Oratio Recta: an Essay on Metarepresentation*. Cambridge, Mass.: MIT Press/Bradford Books.

[3] Faced with the scepticism of his colleagues and friends, Fodor sometimes appeals to an auxiliary argument. He says, or implies, that if we do not accept (CC), we do not explain *why* the constitutive properties of the constituents are inherited by their hosts; we can only *stipulate* that that is so (Fodor 1998b: 53). But I fail to see the force of this argument. The basic fact to be explained is the productivity/systematicity of concepts. To explain that fact, we make two assumptions: Constituency, and Compositionality of reference (CR). We can, if we wish, mention only (CR), since it presupposes Constituency. Be that as it may, once we have Constituency, the *simple* inheritance of constitutive properties is ipso facto explained; it does not have to be stipulated. Nor do we have to enrich (CR) into (CC) in order to explain it. As for *compositional* inheritance, the only difference between Fodor's account, based on (CC), and the alternative account based on (CR), is that Fodor takes *all* constitutive properties of concepts to compose, while the alternative account restricts compositional inheritance to *semantic* properties. I do not see how, without begging the question, one could maintain that one account is more stipulative than the other.

——— (2001). 'Modes of presentation: perceptual vs. deferential'. In A. Newen, U. Nortmann, and R. Stuhlmann-Laeisz (eds.), *Building on Frege: New Essays on Sense, Content, and Concept*. Stanford: CSLI Publications, pp. 197–208.
Stalnaker, R. (1999). *Context and Content*. New York: Oxford University Press.
Strawson, P. (1977). *Logico-Linguistic Papers*. London: Methuen.

11

Keeping Track of Objects in Conversation

Cara Spencer

1.

Suppose Ortcutt and Lingens are talking at a local bar. An eavesdropper overhears two fragments of their conversation. I have numbered some of their utterances for later reference. It is important that the numbers are understood to refer to *utterances* of these sentences in this conversation rather than the sentences themselves.

> Lingens: Have you heard who just got a promotion?
> (1) Ortcutt: In fact, I've just been promoted.
> . . .
> Ortcutt: Who's getting the next round?
> (2) Lingens: Well, you've just been promoted.

The eavesdropper is paying enough attention to each of these fragments to know who is speaking and who is being addressed in each case. But because he has been skulking around the bar between the first and the second fragments, and because he is prone to spatial disorientation, he is unaware that he is eavesdropping on the same table, and the same conversation, twice. As a result, he is unaware that the speaker of (1) is the addressee of (2). His ignorance on this point clearly hinders his appreciation of all that is conveyed here. We typically expect our conversational partners to know when someone is talking about the same object that someone else mentioned previously; that is, to keep track of objects in conversation. The eavesdropper's grasp on this conversation is defective precisely because he has not done so.

I claim that keeping track of objects in conversation is just a matter of having specific beliefs about the object(s) under discussion, beliefs that are typically not literally expressed in conversation. This speaks in favor of treating the propositions that audiences believe when they keep track of objects under discussion as conversational presuppositions. Here, I defend such an account.

In the first part of the paper, I argue that the information the eavesdropper lacks is a part of the conversational background information that audiences expect one another

I presented earlier versions of this paper at the Second Barcelona Workshop on Reference and at the University of Arkansas, and I am grateful to both audiences for helpful discussion that led to substantial changes in this paper. I am also grateful to Lenny Clapp, Michael Glanzberg, David Hunter, Josep Macià, and Robert Stalnaker for very helpful discussion and comments.

to have, without which they could not recover certain kinds of pragmatically conveyed information. For these reasons, I say that they are a kind of speaker presupposition. I then show how Robert Stalnaker's account of assertion content explains how these presuppositions would become a part of the conversational background, and how the audience would use them to recover pragmatically conveyed information.[1] I also propose a two-dimensionalist extension of the basic Stalnakerian account to deal with discourses in which utterances are best understood as conveying the diagonal proposition of a two-dimensional propositional concept. In these discourses, some or all parties to the conversation are confused about exactly which object is being discussed, even though they do keep track of what has been said about it.

2.

I begin with some terminology. If participants in a conversation reasonably expect one another to grasp that two assertions in that conversation are about the same object, then I will say that there is a *discourse-internal identity* between the components of these propositions asserted that pick out this single individual. An interesting question, about which I will have nothing to say here, concerns when exactly we find a discourse-internal identity between assertions in a conversation. I think it is clear that people can talk about the same thing without realizing that they are doing so, and even that people can believe they are talking about the same thing when they in fact are not. I claim only that we find discourse-internal identities in some conversations. Another interesting question, which I have not yet addressed, concerns what proposition we believe when we grasp a discourse-internal identity. I discuss this in Section 3.

What I shall argue here is that if someone understands two contributions to a discourse in isolation but does not grasp the discourse-internal identity between them, then he or she fails to completely understand the discourse, or at least the fragment of it that contains these two utterances. Some may wonder how this is possible. How can someone understand (1) and understand (2), but fail to understand (1) *and* (2)? The answer is that two notions of understanding are at work here, the first from semantics, and the second from pragmatics. I distinguish between understanding (1) and (2), which just involves understanding what (1) and (2) semantically express, and what I will call understanding the discourse (1)–(2), which requires the audience to understand (1) and (2), and further understand that these two utterances are about the same individual. Since I am only concerned here with this specific pragmatic effect, I stipulate that understanding the discourse (1)–(2) does not require us to grasp all of the pragmatic effects of these utterances of (1) and (2), or all background information that audiences use to interpret them.

Clearly the eavesdropper does not get the same information from this discourse as Ortcutt and Lingens do. Why think this information is presupposed? It might instead be background information that Ortcutt and Lingens bring to the conversation, or

[1] The framework I apply here is described in Stalnaker (1978) and (1974).

that they acquire during the conversation, but which has no essential connection to the conversation itself. This is what we might say if we think the eavesdropper understands the conversation perfectly well, but just lacks some information Ortcutt and Lingens happen to have. If, however, the eavesdropper's mistake hinders his understanding of the conversation, then we ought to say that the information he lacks is specifically linguistic, either semantically encoded or pragmatically imparted.

Intuitions about understanding are sensitive to a variety of factors, so they are unlikely to draw a clear distinction between specifically linguistic information and background information. For instance, someone might misunderstand what Quine meant by the dictum, "to be is to be the value of a bound variable," even though he understands what this sentence, or utterances of it, semantically express or pragmatically convey.[2] One may simply lack background information about what is at stake in questions about ontological commitment, and for this reason misconstrue Quine's claim. So we should also look to other considerations to determine how to classify the information the eavesdropper lacks.

There is compelling reason to think this information is not semantically expressed.[3] On a familiar conception, speakers understand what a sentence semantically expresses if they (a) understand the words contained in the sentence, (b) know how the sentence is structured, and (c) know what any indexicals or demonstratives contained in the sentence refer to.[4] On this conception, the information the eavesdropper lacks is not semantically expressed. The eavesdropper, after all, knows (a)–(c) about utterance (1) and utterance (2). It is hardly a settled matter what is involved in knowing who or what an utterance is about. But it cannot require us to know everything about the object in question. Specifically, it cannot require us to know that some other utterance is also about the same individual.[5] So those who accept this conception of semantics are likely to deny that the eavesdropper misses semantically expressed information.

If the information is not semantically expressed, then it is either pragmatically conveyed or a part of the background. We can distinguish a speaker's *private* background information from that which is presupposed by all participants in a conversation. A person's private background beliefs are just his or her beliefs, some of which are relevant to the topic of conversation, and others of which are not. The latter sort

[2] Quine (1948).

[3] An alternative approach to discourse-internal identity is to treat these identities as instances of discourse anaphora. Whether discourse anaphora is a part of the subject matter of pragmatics or semantics is a matter of debate. This approach would involve a radical departure from a widely accepted semantics for indexicals like "you" and "I," according to which the only semantic role for these expressions is to refer to the audience and speaker, respectively, of the context. See Heim (1983) and Kamp (1990) for discussion of this general approach.

[4] Weaker conceptions of semantics, such as that defended by Richard Montague in Montague (1974), consider all effects of context on what a sentence expresses to be the subject matter of pragmatics.

[5] Exactly what is involved in knowing what or who an utterance is about may differ with context, and arguably could include grasp of discourse-internal identities involving it. To go this route to address the problem about discourse-internal identity is to deny that utterances can be understood in isolation.

of background belief, which Robert Stalnaker has called "common ground," is the set of propositions that are mutually believed by all participants in the conversation. That is, they all believe these propositions, and they all take one another to believe them all as well, and they all expect these beliefs to be mutually apparent to everyone participating in the conversation.[6] The common ground, unlike an individual's private background beliefs, plays a role in allowing participants in the conversation to recover pragmatic effects, such as conversational implicature. We can use this fact as the basis for an argument that discourse-internal identities are a kind of speaker presupposition, by showing that audiences who fail to grasp them cannot recover certain pragmatic implicatures.

The first kind of pragmatic implicature that depends on the audience's grasp of discourse-internal identity concerns agreement and disagreement between speakers. Unless they grasp the discourse-internal identities across the conversation, audiences need not appreciate that one speaker has disagreed with what another has said about the object under discussion. Suppose A says "this is F" and B, demonstrating the same thing, says "that is not F." In some such cases, B intends her audience to recognize that she is disagreeing with what A said. To know that A and B are disagreeing when A says "this is F" and B says "that is not F," audiences must be aware that A and B are talking about the same thing. Thus, to grasp what B implicated, that she disagrees with A on this point, audiences must know that A and B are talking about the same thing.

The conversation between Ortcutt and Lingens exemplifies a second kind of pragmatic implicature that also depends on the audience's grasp of discourse-internal identity. Someone who knows that the speaker of (1) is the addressee of (2) knows that these two utterances make the same semantic contribution to the conversation. The semantic content of (1) is old news by the time (2) is uttered, so there is no point in simply reasserting it. Recognizing this, audiences would naturally suppose that Lingens means to pragmatically convey some other proposition, for instance, that since Ortcutt got the raise, Ortcutt should buy the next round. The eavesdropper would be unlikely to grasp this implicature, since he is unaware that these utterances of (1) and (2) have the same semantic content. Since audiences cannot recover either of these pragmatic implicatures unless they grasp the relevant discourse-internal identity, it makes sense to count discourse-internal identities as speaker presuppositions rather than as private background information.

3.

If discourse-internal identities are presupposed and not semantically expressed, we should expect that familiar accounts of the semantic content of utterances do not explain their role. Consider the Russellian theory of content. Russellian theorists argue that the content of a belief is a structured entity whose structure corresponds

[6] See Stalnaker (1974) and (2002).

to that of the sentence that expresses it.[7] Proper names, indexicals and demonstrative pronouns in the sentence contribute only their bearers to this structured content, and other expressions jointly contribute properties or relations to it. Having a belief always involves believing a Russellian content in a certain way, where ways are left largely unspecified save to say that they are associated with sentences. The Russellian would say that the same Russellian content in (3) can be believed in the way associated with either of the sentences in (4) or (5).

(3) <Twain, the property of being a writer>
(4) Mark Twain was a writer.
(5) Samuel Clemens was a writer.

If someone understands a sentence and believes what it says, then according to the Russellian he or she believes the Russellian content the sentence expresses in the way associated with the sentence. So if someone understands and believes an utterance of an indexical sentence and knows who or what the indexicals refer to, then he or she believes a Russellian proposition in the way associated with that indexical sentence.

What does the Russellian account say about our example? Lingens, Ortcutt, and the eavesdropper all understand and believe Ortcutt's initial utterance of (1), so according to the Russellian theory all of them stand in the belief relation to the same Russellian proposition, and they all believe this proposition in the same way. But they may also believe this same proposition in different ways. For instance Lingens' utterance of (2) expresses the same proposition as (1). Lingens, Ortcutt, and the eavesdropper all accept this sentence, so according to the Russellian they also believe this same proposition in another way, associated with (2). Lingens and Ortcutt, unlike the eavesdropper, also believe this proposition in yet another way, associated with a sentence like "You, who produced the first utterance, just got promoted." But since this sentence is not uttered in this conversation, the Russellian theory does not say why an audience would have to have any beliefs in the way associated with it if they are to understand the discourse (1)–(2). The Russellian theory does not explain the special connection between this belief and the conversation.

It is widely assumed that the possible worlds account of content is explanatorily impoverished relative to the Russellian theory.[8] The Russellian theory distinguishes between propositions that necessarily have the same truth value, such as "Bush admires Bush" and "Bush admires himself." And where the Russellian theory distinguishes between a proposition and its logical consequences, the possible worlds view runs afoul of the problem of logical omniscience.[9]

[7] I have in mind the view defended in Salmon (1986) and Soames (1987) and (1995), although the problem I articulate here also arises for John Perry's account of belief (see Perry 1979, 1980a) as he himself has observed (Perry 1980b, 1988).

[8] David Lewis and Robert Stalnaker have defended this account of content. See Lewis (1986) and Stalnaker (1984).

[9] Since the possible worlds view identifies a proposition with a set of worlds, any logical consequence of a proposition will also be true in every world in which the proposition is true. Thus propositions include their logical consequences as a part of their content on this view.

While I agree that the inability to make these distinctions presents a problem for the possible worlds account, the account is also representationally richer than the Russellian theory in certain little-noticed respects. Relevant to this case is its use of the counterpart relation to represent cross-world identities. The counterpart relation is not the identity relation, and admits of a flexibility that the identity relation does not.[10] This flexibility allows us to represent certain states of affairs in the possible worlds framework that cannot be represented in the Russellian framework, at least not without substantial alterations to it.

The argument that the possible worlds approach is more flexible than the Russellian approach is not new.[11] What is new with this paper is the specific application of Stalnaker's dynamic semantics for assertion, combined with the more flexible account of content, to represent what speakers and audiences know when they have kept track of the objects under discussion in a discourse. I use Stalnaker's account of assertion content to represent the information that Ortcutt's utterance (1) and Lingens' utterance (2) are about the same person. Stalnaker's approach represents this piece of information as a presupposition of the discourse fragment that contains (1) and (2), so it provides an explanation of the intuition that Ortcutt and Lingens fully understand the discourse fragment and the eavesdropper does not.

First, some preliminaries about Stalnaker's account of assertion: According to Stalnaker, the content of an assertion, including an assertion made with an indexical sentence, is a set of possible worlds. Assertions are made in the course of a conversation, in which a certain set of presuppositions is operative. This set of presuppositions (or the "context set" as Stalnaker calls it) is what I earlier called the common ground among participants in a conversation. To make an assertion is to narrow down the context set in some way, by excluding some possibilities that remained open prior to the assertion. The content of an assertion, then, is the subset of the context set in which the proposition expressed is true.

Assertions can change the context set in several ways. In the most straightforward case, the speaker rules out those worlds in the context set incompatible with what he has asserted. Information that is not part of the literal content of any assertion can also affect the context set, so long as it is part of the common ground among the participants in the discourse. For instance, the fact that an assertion has been made can affect the context set, since the fact is clearly part of the common ground of participants in the conversation. Contextual facts relevant to interpretation of an utterance will typically, although not always, also be a part of the common ground once the utterance has been made. These include propositions about who is saying what to whom, which things are under discussion, and whether the current speaker is the same one who made a particular earlier contribution to the conversation, and the

[10] Of course, serious metaphysical questions arise about the shape the counterpart relation must take if it is to model the identity relation. Doubtless some philosophers will conclude that a counterpart relation flexible enough to make the distinctions required to account for this and other puzzle cases about belief will not meet the relevant constraints, and for that reason cannot be understood to model the identity relation.

[11] See Stalnaker (1986), (1988).

like. As the conversation progresses, these propositions provide a kind of guide to the roles certain people or objects have played in the conversation. Keeping track of what has been said about whom, who is and has been speaking, and which objects have been under discussion, is part of what is required to understand a conversation. It is precisely this sort of information, information about the context of an utterance that audiences use to interpret it, of which the eavesdropper is unaware. So this account distinguishes what the eavesdropper gets from the conversation and what Ortcutt and Lingens get from it by pointing to a difference in what they presuppose.

On my view, keeping track of objects in conversation is a matter of presupposing propositions involving those objects. Which proposition is presupposed? One way of identifying the proposition is in metalinguistic terms—as a belief that certain occurrences of pronouns co-refer. I think this is the wrong strategy, since it is unlikely that attentive participants in a conversation typically have beliefs of this kind. For one thing, we can understand many conversations, and so keep track of the objects under discussion, without possessing the concepts of reference and co-reference, and without being able to form metalinguistic beliefs. Furthermore, even if we have these concepts, we can keep track of what has been said about a single object even after we have lost track of the specific utterances used to refer to it. So the beliefs in question are not beliefs about the words, or tokens of those words, used to refer to an object over the course of a conversation. Rather, when we keep track of an object under discussion, we presuppose singular propositions about that object, and we make different presuppositions about it at different times in the conversation.

Specifically, the presupposed singular proposition is expressible with an identity statement. This identity statement is not uttered in the conversation, but some of its components are. These components are the singular referring terms in the utterances involved in the discourse-internal identity itself. So for instance, they are the occurrences of "I" in (1) and "you" in (2). The singular proposition at issue, then, contains just those worlds in which the referents of these utterances of these terms co-refer. How is this proposition different from the above-mentioned metalinguistic proposition, which is about these two utterances? First, the metalinguistic proposition is about these two tokens of "I" and "you," and the singular proposition at issue is not about them. This explains why audiences would have to keep track of specific utterances about an object, or their parts, to grasp the metalinguistic proposition. It seems more likely that when we keep track of objects in conversation, we do so by recalling the content of what is said or conveyed in a conversation, not by recalling the utterances themselves. Second, since there are some worlds in which those utterances of "I" and "you" do not refer to their actual referents, the metalinguistic proposition and the singular proposition simply differ in content.

If these two utterances of referring terms ("I" in (1) and "you" in (2)) actually co-refer, then an apparent problem for my account arises immediately. Since identities are necessary, this singular identity proposition is also necessary. It is true in all possible worlds, thus it is presupposed whether or not there is a discourse-internal identity between these two utterances. There are really two problems here. The first is specific to the possible worlds account of content, according to which all necessary propositions (or strictly *the* necessary proposition, which is just the set of all possible

worlds) are true in any set of possible worlds. The second is a more general and much-discussed problem about the informativeness of identity statements. Stalnaker has offered an account of identity statements that addresses both problems as they arise here. The solution uses Stalnaker's notion of a two-dimensional matrix associated with an utterance, so let me first turn to that notion.

Specifically semantic context sensitivity ensures that a single utterance will express different propositions in different possible worlds. For instance, an utterance of "I am sitting" expresses the proposition that Fred is sitting in a world where Fred produces that utterance, and it expresses the proposition that Jane is sitting in a world where Jane produces it. Suppose that Fred is sitting in worlds w and w*, and Jane is sitting only in world w*, and that Jane produces the utterance of "I am sitting" in w*, and Fred produces it in w. The figure below represents a two-dimensional propositional matrix for world w and w* that is determined by this utterance of "I am sitting."

	w	w*
w	T	T
w*	F	T

Along the top row of the matrix, worlds w and w* are considered to be worlds of evaluation, and along the left side, they are considered as worlds of utterance. In general, the cell in the *i*th row and the *j*th column of the matrix gives the semantic value of the utterance considered as uttered in world w^i and evaluated in w^j.[12] The horizontal proposition associated with an utterance is just the set of worlds *w* in which the utterance, as uttered in the actual world, is true in *w*. This is what we would normally think of as the proposition expressed by the utterance. The two-dimensional matrix determines many other propositions, one of which is the diagonal proposition, which is the set of worlds *w* such that what *u* expresses in *w* is true in *w*. In the example, this set contains both *w* and *w**, as is shown. On Stalnaker's view, the content of an assertion is always a subset of the context set for the conversation in which the assertion is made. The same goes for the diagonal proposition. So for Stalnaker's purposes, the relevant two-dimensional matrix for an utterance will not include all possible worlds, but only those worlds contained in the context set for that utterance. Since beliefs about the meanings of words used in the conversation are a part of the common ground for a conversation, only those worlds in which these beliefs are true will be included in the context set.

Stalnaker proposes that the diagonal proposition and the horizontal proposition associated with an utterance can be candidates for the proposition it expresses. The availability of the diagonal proposition for this purpose affords an explanation of the informativeness of identity statements. If an identity statement is informative for an audience, then that audience does not presuppose that the two terms flanking the identity sign co-refer. It would be natural to think that the context set for an

[12] This generalization holds only if the matrix is constructed with *n* worlds labeled $w–w^n$ arranged in numerical order in both the top row and the leftmost column.

informative identity statement includes worlds in which they do co-refer, and worlds in which they do not. The horizontal proposition that an identity statement expresses is still a necessary truth, but the diagonal proposition will not be.

Why think that Ortcutt and Lingens presuppose this diagonal proposition when (2) occurs, but the eavesdropper does not? To answer this question, we need to consider the open possibilities for Ortcutt and Lingens, and compare them with the open possibilities for the eavesdropper. The eavesdropper does not believe that the speaker of (1) is the addressee of (2). So the eavesdropper has two representations of Ortcutt, both of which are sensitive to information that the eavesdropper gets through causal interaction with the real Ortcutt. Since Ortcutt and Lingens are aware that the addressee of (2) is the same individual who said earlier that he was promoted, they only have a single representation of Ortcutt. Since some of the eavesdropper's belief worlds contain two individuals, one associated with (1) and another associated with (2), the diagonal proposition at issue is not true in all of his belief worlds.

How can the possible worlds view accommodate the claim that the eavesdropper has two representations of a single actual individual? There is only one actual Ortcutt, and the eavesdropper has two singular representations of him. To say that the representations are singular is to say that they would not exist if Ortcutt did not. On the possible worlds view, individuals and their counterparts in other possible worlds are used to represent the content of a singular belief. So for instance, the content of the belief that Salvador Dalí was a painter is the set of worlds in which Dalí, or his counterparts in other possible worlds, are painters. If the eavesdropper has two representations of Ortcutt, and reserves judgment about whether they are representations of the same individual, then some possible worlds in the eavesdropper's belief set contain two individuals, both of which are Ortcutt's counterparts. One of these individuals is associated with (1) and the other with (2). Thus the diagonal proposition at issue is false in these worlds.

At this point a metaphysical problem arises. What would it mean to say, as I do, that there are some worlds in the eavesdropper's belief set in which Ortcutt has two counterparts? Is that not just saying that there could have been two Ortcutts, and is that not impossible? There are several ways out of this metaphysical problem. We might follow Robert Stalnaker in arguing that there are worlds in which one actual object has more than one counterpart. On Stalnaker's view, each possible world has its own domain of individuals, and no individual exists in more than one possible world. The only world that contains Ortcutt is the actual world, and nothing in any other possible worlds is identical to him. We can nonetheless interpret claims that an actual individual might have had property P by designating an individual that has property P in the domain of another possible world to serve as the counterpart to the actual individual. Of course, we cannot just designate any individual we choose to serve as this counterpart, not if the counterpart relation is to model the identity relation. The counterpart relation will have to be constrained by a general metaphysics for objects, and it should have certain formal properties as well. For instance, the counterpart relation should be symmetric and transitive like the identity relation. Stalnaker has suggested that a counterpart relation that meets these formal constraints can still

allow that one actual object can have more than one counterpart in some possible worlds.[13] Another way out, also due to Stalnaker, is to accept that identities are necessary and hold that it is indeterminate which of the two individuals in the eavesdropper's belief worlds is Ortcutt's counterpart. That is, we might hold that there are two sets of possible worlds in which there are two individuals who have just been promoted. For all the eavesdropper knows, one of these individuals is the subject of the first utterance (1), and the other is the subject of the second utterance (2). In one of these sets, the first individual is Ortcutt's counterpart, and in the other, the second individual is Ortcutt's counterpart. We might then say that it is indeterminate which of these two sets of possible worlds is the content of the eavesdropper's beliefs.

One might of course reject the view that one actual object can be two or more objects in another possible world, or that what someone believes may be to some degree indeterminate. If neither position is correct, then the Russellian and possible worlds accounts of content will have the same expressive resources with respect to this example. Both positions are substantive claims about the metaphysics of identity in the first case and the nature of belief in the second, so it is false that the possible worlds account of content only makes distinctions already available to the Russellian. If either of these positions can be made plausible, then the possible worlds approach distinguishes between the cognitive situations of Ortcutt and Lingens, who understand the conversation, and the eavesdropper, who does not.[14]

4.

We keep track of the object under discussion when it is common knowledge which thing is being tracked. But we also do this when it is not. Suppose, for instance, that Ortcutt dials his friend O'Leary. O'Leary lives with his son, whose voice is indistinguishable from his over the telephone. Someone answers the phone, and Ortcutt can tell immediately that it is either O'Leary or his son, but cannot tell which one it is. Embarrassed to admit his ignorance, Ortcutt simply continues the conversation, hoping that some conversational clues will fill him in on which of the two he is actually talking to. Unfortunately for Ortcutt, the conversation continues for some time before he realizes who is on the other end of the line.

I want to make a few points about the example. First, Ortcutt can use referring terms, such as "you," to refer to the person he is talking to. Second, he can keep track of what has been said about this individual, and even about other individuals mentioned during the conversation, without knowing which of several individuals he is keeping track of.[15] Third, this case is importantly different from the more commonly

[13] See Stalnaker (1986), (1988).

[14] Although Stalnaker's account of speaker presupposition uses the possible worlds account of propositions, one might accept the former account and reject the latter. I have not considered how one might combine the Stalnakerian treatment of discourse-internal identity with other accounts of the proposition, but I see no reason in principle that such accounts could not be given.

[15] We can imagine, for instance, that the person on the other end of the line starts talking about his sister's recent accomplishments. Since both O'Leary and his son have sisters, Ortcutt is still in

discussed possibility in which we refer to an object, and even have singular thoughts about it, without being able to distinguish it from other similar objects.

I think the first two points are uncontroversial, but I will say a little in defense of the third point. Suppose Fred receives a call from a telemarketer. Fred can have singular thoughts about the telemarketer on the other end of the line even if he has no clear idea about who he or she is.[16] The difference between this typical case of singular thought and Ortcutt's beliefs about the person on the other end of the line is that in the latter case, Ortcutt has in mind two individuals such that he knows that one of them is the person he is speaking to but he does not know which one it is. There are two possibilities, two candidates for the person Ortcutt's thoughts are about, that are relevant to characterizing his beliefs. Unlike Ortcutt, Fred will be unable to distinguish many individuals from the one his thought is about, but we can characterize the content of Fred's belief without referring to these other objects at all. Since Ortcutt knows that he is speaking to either O'Leary or his son, the possible worlds account of belief content characterizes Ortcutt's belief, say, that the person on the other end of the line just bought a new car, as the set of worlds in which either O'Leary (or his counterpart) just bought a new car or O'Leary's actual son (or his counterparts) just bought a new car. Fred's singular belief about the telemarketer, say, that she is trying to sell him an air-conditioner, is not about different individuals in different possible worlds. The content of this belief, on the possible worlds view, is the set of worlds in which the actual telemarketer is trying to sell Fred an air-conditioner. Fred's relative dearth of descriptive information about the telemarketer is reflected in the fact that the actual telemarketer has many different properties in the possible worlds in Fred's belief set. In some worlds, she is a middle-aged woman with red hair in Milwaukee, in others a woman with dark hair calling from Omaha. Each of these individuals is a world-bound counterpart to the actual telemarketer Fred is speaking to. The reason for treating the cases differently has to do with the differences in the believers' cognitive states. Ortcutt's representations of O'Leary and his son are both parts of the content of his thought, but Fred only has one representation of the telemarketer that is relevant to the content of his thought.

One might be reluctant to say that two individuals, O'Leary and his son, should be used to characterize the content of Ortcutt's belief. A powerful reason for such reluctance rests on the intuition that the Ortcutt example could not be different from the telemarketer example. Ortcutt and Fred both stand in the same kind of causal-informational connection to the person on the other end of the phone line. If this connection suffices to make Fred's belief a singular belief about a single individual, why does it not suffice in Ortcutt's case? I think the intuition is misguided. Ortcutt's causal relation to the person on the other end of the line does not consist solely of

the dark about who he is talking to. Still, he can keep track of what is said about the sister, even though he does not know whose sister he is talking about.

[16] Gareth Evans suggested that if a thinker is connected to an object via an information link of the sort that makes perception of that object possible, the agent can have singular thoughts about that object (Evans 1982). I assume that hearing a person's voice over a phone line is a suitable information link.

the causal-informational link the telephone provides. Ortcutt has two rich singular concepts, one of O'Leary, and the other of his son, and Ortcutt's beliefs about the person on the other end of the line clearly invoke both singular concepts.

The interest of this sort of case is that it demonstrates the independence of keeping track of an object through a conversation and knowing which object you are keeping track of. If the two are independent, then we should expect that an account of keeping track of objects in conversation should be applicable to this sort of case as well. A virtue of the account I defend is that it can be extended in this way. In the example I consider below, the conversational common ground does not specify which of two objects the conversation is about, yet participants in the conversation nonetheless keep track of it.

> Watergate: Washington Post reporters Bob Woodward and Carl Bernstein receive important leaks about the break-in at the Watergate Hotel from a source they identify only as "Deep Throat." Suppose, improbably, that a Post intern, Sue, has been assigned to take messages from Deep Throat. She has determined through her own investigation that the man she calls "Deep Throat" is one of two people, either W. Mark Felt or Fred Fielding, and she already has told as much to Deep Throat. They have the following telephone conversation:
>
> Sue: The last time we spoke, you were calling from your office.
> Deep Throat: At that point, you had almost figured out who I am.

What does the approach introduced above say about this case? Sue knows that her conversational partner is either Felt or Fielding, but does not know which one he is. Deep Throat knows that she knows this, and since she is the one who told him so, Sue knows that he knows this. So when Sue speaks, it is common ground that her addressee is either Felt or Fielding, but the common ground does not specify which one of the two he is. Thus, the context set will contain worlds in which her addressee is Fielding, and worlds in which he is Felt. Since the addressee of Sue's utterance differs from world to world in the context set, the utterance does not express the same proposition in each of these worlds.

Stalnaker has suggested that one of what he calls the "essential principles of rational communication" is that an utterance should express the same proposition relative to every world in the context set.[17] An assertion is supposed to narrow down the context set in some way or other, and if an utterance provides a different "instruction" for narrowing the context set in different worlds, then audiences will not know which instruction to follow. Since Sue's utterance of "you" refers to Fielding in some worlds in the context set, and to Felt in others, this conversation appears to violate this principle. Stalnaker has suggested that when conversations violate the essential principles he articulates, participants may deal with the violation by interpreting the utterance differently, so that it expresses the same proposition in every world. The diagonal proposition associated with the utterance is a natural candidate interpretation.

[17] Stalnaker (1978), 88.

Let us suppose, then, that the content of Sue's utterance is the associated diagonal proposition rather than the horizontal. This is the set of worlds in which Sue's addressee was calling from his office the last time they spoke. Specifically, it is the set of worlds in which either Fielding was calling from his office the last time they spoke and Fielding is the addressee of Sue's utterance, or Felt was calling from his office the last time they spoke and Felt is the addressee of Sue's utterance. Deep Throat's utterance, like Sue's, expresses different propositions in different worlds in the context set, since in some of these worlds the speaker is Felt and in others it is Fielding. Here again we can suppose that the diagonal proposition best represents the content of Deep Throat's utterance. This will be the set of worlds *w* such that either Felt is the speaker in *w* and Sue had almost figured that out the last time they spoke, or Fielding is the speaker in *w* and Sue had almost figured that out the last time they spoke.

In this conversation, there is a discourse-internal identity between the occurrences of "you" in Sue's utterance and "I" in Deep Throat's utterance. What effect does this presupposition have on the context set? Consider what worlds would remain in the context set without it. In some of these worlds, the addressee of Sue's utterance is not the speaker of the next utterance. But both parties to the conversation know that these possibilities should be excluded. When Deep Throat speaks, the propositions that he is speaking, and that he was the addressee of Sue's last utterance, become part of the context set. Thus it is presupposed that the speaker of the second utterance is the same as the addressee of the first. And it is this presupposition, not the content of any assertion, that rules out these possibilities. Hence we have the desired result, that the referents of Sue's "you" and Deep Throat's "I" differ from world to world in the diagonal proposition, but in each world they are the same.

It is a commonplace of pragmatics that audiences must know how a discourse hangs together if they are to grasp all of the pragmatic effects of any specific utterance in that discourse. One way a discourse can hang together is by being about a single object. Keeping track of which utterances are about the same object involves constantly updating a store of information as the conversation progresses. Stalnaker's dynamic account of assertion content is for this reason an ideal vehicle for representing the changing set of presuppositions that audiences use to interpret each utterance as it occurs. If, as I have suggested, audiences can keep track of objects in conversation without knowing exactly which objects are being discussed, then the Stalnakerian framework has the virtue of treating both kinds of keeping track as instances of the same phenomenon.

References

Evans, Gareth (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Heim, Irene R. (1983). "File change semantics and the familiarity theory of definiteness" in Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, eds., *Meaning, use, and interpretation of language*. Berlin: de Gruyter.

Kamp, Hans (1990). "A Prolegomena to a Structural Account of Belief and Other Attitudes" in C. Anthony Anderson and Joseph Owens, eds. *Propositional Attitudes*. CSLI Lecture Notes, no. 20. Stanford, CA: CSLI.

Lewis, David (1986). *On the Plurality of Worlds*. Oxford, England and New York: Basil Blackwell.

Montague, Richard (1974). "Pragmatics" in R. Thomason, ed. *Formal Philosophy*. New Haven: Yale University Press.

Perry, John (1979). "The Problem of the Essential Indexical." *Noûs* 3: 3–21.

____(1980a). "Belief and Acceptance." *Midwest Studies in Philosophy* 5: 533–4.

____(1980b). "A Problem about Continued Belief." *Pacific Philosophical Quarterly* 61: 317–32.

____(1988). "Cognitive Significance and New Theories of Reference." *Noûs* 22: 1–18.

Quine, W. V. (1948). "On What There Is." *Review of Metaphysics* 2: 21–38.

Salmon, Nathan (1986). *Frege's Puzzle*. Cambridge, Mass.: MIT Press.

Soames, Scott (1987). "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15: 47–87.

____(1995). "Why Singular Propositions?" *Canadian Journal of Philosophy* 25: 515–50.

Stalnaker, Robert (1974). "Pragmatic Presuppositions" in M. K. Munitz and P. Unger, eds. *Semantics and Philosophy*. New York: New York University Press.

____(1978). "Assertion." *Syntax and Semantics* 9. New York: Academic Press. Reprinted in Stalnaker (1999). Page references from this reprint.

____(1984). *Inquiry*. Cambridge, Mass.: MIT Press.

____(1986). "Counterparts and Identity," in P. A. French, T. E. Uehling, and H. K. Wettstein, eds. *Midwest Studies in Philosophy* 9, Minneapolis: University of Minnesota, 121–40.

____(1988). "Belief Attribution and Context" in R. Grimm and D. Merrill (eds.). *Contents of Thought*. Tucson: University of Arizona Press.

____(1999) *Context and Content*. New York: Oxford University Press.

____(forthcoming). "Common Ground." *Linguistics and Philosophy* 25, 5–6: 701–21.

12

Kripke, the Necessary Aposteriori, and the Two-Dimensionalist Heresy

Scott Soames

The Necessary Aposteriori: Revolution and Reaction

A little over thirty years ago, Saul Kripke and others advanced what was then a revolutionary thesis: there are necessary truths knowledge of which requires empirical evidence. Kripke's route to this conclusion was breathtakingly simple. He first used the concept of rigid designation to rebut Quine's influential objection to essentialism.[1] Then, with both a non-descriptive semantics and a rehabilitated conception of essentialism in place, he showed how to generate instances of the necessary aposteriori. If n is a rigid designator of o, and P expresses an essential property of o which is such that knowledge that o has it requires empirical evidence, then the proposition expressed by ***if n exists, then n is P*** is both necessary and knowable only aposteriori.[2]

Although Kripke's examples were extraordinarily convincing, some theorists harbored philosophical commitments that did not allow them to be convinced. For example, those who identified propositions with sets of metaphysically possible world-states were committed to the view that there is only one necessary proposition, which surely is knowable apriori.[3] Other theorists offered analyses of knowledge according to which one knows p iff one has evidence ruling out all relevant (metaphysically) possible ways in which p could be false—a conception according to which necessary truths are trivially knowable, since they are true no matter which possible state the world is in.[4] For these theorists, the only option was to try to explain away Kripke's revolutionary discovery. In the last twenty-five years a systematic strategy has grown up around a technical development called *two-dimensional modal logic* for doing essentially that. This strategy seeks to construct a descriptivist two-dimensional model that lays the foundation either for denying

Thanks to Ali Kazmi and Ben Caplan for useful comments and discussion.

[1] See chapter 14 of Soames (2003).

[2] Here, and throughout, I use boldface italics to play the role of corner quotes. Ordinary bold is used for emphasis, ordinary italics to mention expressions.

[3] See Robert Stalnaker, "Assertion," originally published in 1979, reprinted in Stalnaker (1999); also see Stalnaker (1984).

[4] See Lewis (1996), reprinted in Lewis (1999).

the necessary aposteriori altogether, or for draining it of much of its philosophical significance. In what follows, I will say a few words about the origins and defects of this strategy.

An Objection to the Necessary Aposteriori

I begin with a puzzle about the idea that any single proposition p can be both necessary and knowable only aposteriori. To say that p is knowable only aposteriori is to say that empirical evidence supporting its truth is required in order to justify one's belief in p. But how, if p is necessary, can empirical evidence about the actual world-state be required to establish p? Surely, if evidence is required, it must have the function of ruling out possible ways in which p could be false. But, if p is true no matter which possible state the world is in, then there are no such ways. So, if p really is necessary, there should be no need for evidence justifying it, in which case p should be knowable apriori, if it is knowable at all. On the other hand, if p really is knowable aposteriori, then there must be genuinely possible ways that the world could be in which p is false. Hence, the necessary aposteriori is impossible. David Lewis argues this way in his recent paper, "Elusive Knowledge." However, the problem is not new. There is even a suggestion of it in *Naming and Necessity*.

Kripke's Response to this Objection in *Naming and Necessity*

Kripke's statement of the problem and sketch of a solution

In the middle of lecture 3, after summing up his treatment of natural kind terms and illustrating their role in generating examples of the necessary aposteriori, Kripke takes up a challenge to his view. Up to this point, when discussing necessary aposteriori truths, he has emphasized that although they are necessary, and hence true with respect to every possible world-state, nevertheless, for all we knew prior to empirically discovering their truth, they, in his words, "could have turned out otherwise." Realizing that this may sound puzzling, he gives voice to the following objection.

> Theoretical identities, according to the conception I advocate, are generally identities involving two rigid designators, and therefore are examples of the necessary *a posteriori*. Now in spite of the arguments I gave before for the distinction between necessary and *a priori* truth, the notion of *a posteriori* necessary truth may still be somewhat puzzling. Someone may well be inclined to argue as follows: 'You have admitted that heat might have turned out not to have been molecular motion, and that gold might have turned out not to have been the element with the atomic number 79. For that matter, you also have acknowledged that . . . this table might have turned out to be made from ice from water from the Thames. I gather that Hesperus might have turned out not to be Phosphorus. What then can you mean when you say that such eventualities are impossible? If Hesperus might have *turned out* not to be Phosphorus, then Hesperus might not have *been* Phosphorus. And similarly for the other cases: if the world could have *turned out* otherwise, it could have *been* otherwise.[5]

[5] Kripke (1980), pp. 140–1.

The problem here starts out being about theoretical identity sentences involving natural kind terms, but quickly expands to cover all instances of the necessary aposteriori. Let p be such an instance. Since p is aposteriori, its falsity must be conceivable, and we need empirical evidence to rule that out. Without such evidence **it could turn out that** p is false. But, the objector maintains, if p is necessary, there are no such possibilities. So, if p really is necessary, we do not require empirical evidence to know p after all; and if p really is aposteriori, then p is not necessary. The necessary aposteriori is an illusion.

Kripke begins his reply to this objection with the following passage.

> The objector is correct when he argues that if I hold that this table could not have been made of ice, then I must also hold that it could not have turned out to be made of ice; *it could have turned out that P* entails that P could have been the case. What, then, does the intuition that the table might have turned out to have been made of ice or of anything else, that it might even have turned out not to be made of molecules, amount to? I think that it means simply that there might have been *a table* looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. In other words, I (or some conscious being) could have been *qualitatively in the same epistemic situation* that in fact obtains, I could have the same sensory experience that I in fact have, about *a table* which was made of ice.[6]

Imagine the following scenario: a table has been brought in, I have examined it and determined it to be made out of wood, not ice. I point to the table and say *I know that this table is not made out of ice*. I know this because I have empirically ruled out what otherwise would have been an epistemologically relevant possibility. Prior to my checking, **it could have turned out**, for all I knew, that the table was made of ice. The intuition that things could have turned out that way is, Kripke suggests, nothing more than the judgment that it is genuinely possible for me, or some other agent, to be in a situation qualitatively identical to this one, and be pointing at a table that **is** made out of ice.[7]

He generalizes this point in the next paragraph.

> The general answer to the objector can be stated, then, as follows: Any necessary truth, whether *a priori* or *a posteriori*, could not have turned out otherwise. In the case of some necessary *a posteriori* truths, however, we can say that under appropriate qualitatively identical evidential situations, **an appropriate corresponding qualitative statement** might have been false. The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that there should have been a compound with all the properties originally known to hold of gold. The inaccurate statement that Hesperus might have turned out not to be Phosphorus should be replaced by the true contingency mentioned earlier in these lectures: two distinct bodies might have occupied, in

[6] Kripke (1980), pp. 141–2.

[7] The suggestion is, as we shall see, problematic. However, even at this stage there is something surprising about the application of this idea to the intuition that the table might have turned out not to be made of molecules. Is Kripke suggesting that it is genuinely metaphysically possible that **some table** might not be made out of molecules? One would have thought that the claim that physical objects like tables are made up of molecules would count as a metaphysically necessary truth, on a par with the claim that water is made up of molecules that contain two hydrogen atoms and one oxygen atom.

the morning and the evening, respectively, the very positions actually occupied by Hesperus-Phosphorus-Venus.[8]

This paragraph and the one preceding it mark the beginning of what, in my opinion, is the most misleading and potentially problematic passage in *Naming and Necessity*.

Two main issues are addressed: the necessity of certain propositions and the fact that they can be known only aposteriori. Regarding the former, Kripke makes three points:

(i) There is a natural and correct way of understanding the locution ***it could have turned out that*** $\sim S$ in which it entails ***it is not necessary that S***.

(ii) When understood in this way, his previous remarks—that when S is both necessary and aposteriori, empirical evidence is needed because ***it could have turned out that*** $\sim S$ was true—were strictly speaking inaccurate.

(iii) In these cases the necessary proposition expressed by S is easily confused with certain descriptive propositions that are both contingent and knowable only aposteriori. These are the propositions that could genuinely have turned out not to be true.

Kripke maintains that when the objector protests that his examples cannot be necessary, given that they are aposteriori, the objector is **confusing** the propositions expressed by the examples with other, related propositions that really are contingent. The objector confuses the singular proposition that **this** table in front of me is made out of ice with the related general proposition that the, or a, table in front of me is made out of ice. He also confuses the necessary truth expressed by (1a) with the contingent truths expressed by (1b–c).

1a. Hesperus is Phosphorus
b. 'Hesperus' and 'Phosphorus' are coreferential.
c. 'Hesperus is Phosphorus' expresses a truth in our language.

Since the two names are associated with a pair of descriptions that cannot be satisfied unless the heavenly body that appears in the evening sky is the heavenly body that appears in the morning sky, the objector also ends up confusing the necessary truth expressed by (1a) with the contingent truth expressed by (1d).

1d. The heavenly body that appears in the evening sky (at time t and place p) is the heavenly body that appears in the morning sky (at t^* and p^*).

This response of Kripke's is unobjectionable, as far as it goes. However, it does not go far enough. Although it deals with objectors who grant that his examples are aposteriori, but doubt they are necessary, it does not deal with objectors who grant that the examples are necessary, but doubt that they are aposteriori. More importantly, the reply fails to deal with the general form of the objection, which purports to demonstrate, without relying on any particular example, that no proposition can be both

[8] Kripke (1980), pp. 142–3, my emphasis.

necessary and knowable only aposteriori. Moreover, to the extent that his remarks do suggest a reply to these worries, it is puzzling and inadequate. In the passage, Kripke seems to suggest that his earlier argument that the claim that Hesperus is Phosphorus is knowable only aposteriori provides the pattern of explanation for all other examples of the necessary aposteriori. This is unfortunate.

Kripke's argument that it is not knowable apriori that Hesperus is Phosphorus

Kripke's argument that it is not knowable apriori that Hesperus is Phosphorus, given in the last four pages of lecture 2, is based on the observation that evidence available to us simply by virtue of understanding the names *Hesperus* and *Phosphorus* is insufficient to show that they are coreferential. Since agents in epistemological situations qualitatively identical with ours might use the names exactly as we do, yet be referring to different things, the qualitative evidence we have by virtue of understanding the names is insufficient to justify the claim that they are coreferential. Thus, the metalinguistic claims (1b) and (1c), as well as the non-metalinguistic claim (1d), are not knowable apriori. This is, of course, correct. However, it is not the conclusion Kripke is interested in. The conclusion he explicitly draws is that it is not knowable apriori that Hesperus is Phosphorus.[9] Unfortunately, this non-metalinguistic conclusion does **not** follow from his stated premises. The proposition that Hesperus is Phosphorus is, as he insists, true in all possible world-states. So it is true in all world-states in which agents are in epistemic situations qualitatively identical to ours. Hence, the principle that only propositions true in all such states are knowable apriori does not rule out that it may be knowable apriori.[10]

Perhaps, however, the gap in Kripke's argument can be filled. Throughout the passage, he exploits a familiar connection between speakers' understanding and acceptance of sentences and our ability to use those sentences to report what they believe. In his example, before we learned of the astronomical discovery, we understood but did not accept (1a); hence it is natural to conclude that we did not believe that Hesperus was Phosphorus. Moreover, we would not have been **justified** in accepting (1a) based on the evidence we had at that time. Because of this, it is natural to think that we would not have then been justified in **believing** that Hesperus is Phosphorus. If so, then the proposition that Hesperus is Phosphorus must require empirical justification, in which case it must not be knowable apriori—exactly as Kripke says.

With this in mind, we may reconstruct Kripke's implicit reasoning as follows.

(i) One who understands *Hesperus is Phosphorus* accepts it and believes it to be true iff one believes that Hesperus is Phosphorus.

[9] See pp. 103–4 of Kripke (1980). Although in the passage on pp. 142–3 quoted above Kripke mentions the non-metalinguistic (1d) without mentioning the metalinguistic (1b) and (1c), he implicitly refers the reader to the passage on pages 103–4, where he mentions both types of examples, and concentrates on the metalinguistic.

[10] For further discussion, see Soames (2002), pp. 6–9.

(ii) Similarly, one who understands *Hesperus is Phosphorus* would be justified in accepting it and believing it to be true iff one would be justified in believing that Hesperus is Phosphorus.

(iii) In order to be justified in accepting *Hesperus is Phosphorus* and believing it to be true, it is not sufficient for one simply to understand it; in addition one needs empirical evidence that the two names refer to the same thing.

(iv) Therefore, understanding *Hesperus is Phosphorus* is not sufficient for one to be justified in believing that Hesperus is Phosphorus; in addition, one must have empirical evidence that the two names refer to the same thing.

(v) Therefore the statement that Hesperus is Phosphorus is not knowable apriori.

This, I take it, is the reasoning Kripke uses to support his conclusion that the necessary truth that Hesperus is Phosphorus is knowable only aposteriori, and it is the reasoning that he seeks to generalize to other cases of the necessary aposteriori. The key elements in the reasoning are the principles of Strong Disquotation and Strong Disquotation and Justification, which, without fussing over details, may be formulated roughly as follows:

> *Strong Disquotation*
> If x understands S, uses it to express p, and knows that S expresses p, then x believes p iff x accepts S (and believes it to be true).
> *Strong Disquotation and Justification*
> If x understands S, uses it to express p, and knows that S expresses p, then x would be justified in believing p on the basis of evidence e iff x would be justified in accepting S (and believing it to be true) on the basis of e.

How are these principles used? If I understand the sentence (1a), *Hesperus is Phosphorus*, while associating the two names with the descriptions *the heavenly body seen in the evening sky (at t and p)*, and *the heavenly body seen in the morning sky (at t* and p*)*, then I will justifiably accept (1a) **only if** I justifiably **believe** that the heavenly body seen in the evening sky (at t and p) is the heavenly body seen in the morning sky (at t* and p*). Since my justification for this descriptive belief is empirical, my justification for accepting sentence (1a) is also empirical. Strong Disquotation and Justification will then tell us that my belief in the proposition I use the sentence to express—presumably the proposition that Hesperus is Phosphorus—is empirically justified. Hence, my knowledge of this proposition is aposteriori. If one assumes that this result carries over to other agents, times, and sentences expressing the same proposition, then one will arrive at Kripke's conclusion that this proposition can be known **only** aposteriori.

Next consider the table that has been brought into the room. In pointing at it and saying *This table is not made out of ice*, I express a necessary truth—since **this very table** could not have been made out of ice. Nevertheless, in this context I would not accept, and would not be justified in accepting, the sentence *This table (pointing) is not made out of ice* unless I **also** believed, and was **justified** in believing, the general descriptive proposition that the, or a, table directly in front of me is not made out of ice. This descriptive proposition q is, of course, contingent rather than necessary, and

hence not to be confused with the proposition expressed by the indexical sentence I uttered. Since I am justified in believing q only on the basis of empirical evidence, and since this evidence is **included** in the evidence on which I base my utterance, my evidence for accepting the sentence uttered must also be empirical. From strong disquotation and justification, it follows that although it is a necessary truth that this table is not made out of ice, my knowledge of this truth is based on empirical evidence, and so is aposteriori. Generalizing to other agents, times, and ways of expressing the same proposition, one might well conclude that this proposition is both necessary and knowable **only** aposteriori.

These examples illustrate Kripke's strategy for answering the objection to the necessary aposteriori. Confronted with someone who grants that S expresses a necessary proposition p, but objects that since p is necessary, knowledge of it cannot require empirical justification, Kripke replies that empirical evidence is required in order to know a **different** but qualitatively similar proposition q that is related to p in a certain way. When he speaks of S as being something that "could have turned out false," and hence requires empirical justification, he has in mind not the proposition p actually expressed by S, but a corresponding qualitative proposition q that **is** false in certain possible world-states involving agents in epistemic situations qualitatively identical to ours. This contingent proposition is one the agent must know in order to be counted as knowing the necessary proposition p expressed by S.

The structure of Kripke's response

That is Kripke's final, problematic response to the objector. Recall the objector's argument. If p is knowable only aposteriori, then empirical evidence is needed to rule out certain possible circumstances in which p is false. But, if p is necessary, there are no such circumstances to rule out. Thus, no proposition can be both necessary and knowable only aposteriori. To this, two main replies could be made; one could reject either P1 or P2.

P1. When empirical evidence is required for knowledge of p its function is to rule out possibilities in which p is false.

P2. All epistemic possibilities are genuine, metaphysical possibilities—roughly, every way that, for all we know apriori, the world might be is a way that the world genuinely could be.

One would have thought that Kripke was committed to rejecting P2 anyway, in which case nothing more would need to be said to rebut the objector's argument.[11] What we have seen, however, is that these few pages of Kripke's text can be read as suggesting something quite different—namely, the rejection of P1, and its replacement by P3 and P4.

[11] Kripke's discussion of Goldbach's conjecture, p. 35 ff, indicates that he does not rule out epistemic possibilities that are not metaphysical possibilities. In the final section below, I explain why he should be understood as embracing them.

P3. When empirical evidence is required (by the agent) for the truth of ***a knows that S***, its function is always to rule out possibilities. However, sometimes the possibilities to be ruled out are **not** those in which the proposition expressed by S is false; instead they are possibilities in which a certain related proposition is false.

P4. Examples of the necessary aposteriori are those in which even though S expresses a necessary truth p, the truth of ***a knows that S*** always requires knowing some contingent, aposteriori proposition q that is related to p in a certain way.

Since it is P3 that is most objectionable, I will not here worry about P4. In what follows, I will relate Kripke's route to P3 to a prototypical version of two-dimensionalism, and briefly indicate why both views are incorrect. I will then finish up by sketching a natural Kripkean strategy that rebuts the objector's argument against the necessary aposteriori by rejecting P2 rather than P1.

Heresies

Kripke's Strong Disquotational route to P3

The problem with principles of strong disquotation is that they require an unrealistic degree of transparency of meaning. Sentences S_1 and S_2 may mean the same thing, and express the same proposition p, even though a competent speaker who understands both sentences, and associates them with p, does not realize that they express the same proposition. Such an agent may accept S_1, and believe it to be true, while refusing to accept S_2, or to believe it to be true. This is the situation that Kripke's well-known character Pierre finds himself in with the sentences *Londres est jolie* and *London is pretty*.[12] Although both mean that London is pretty, and although Pierre understands both, he does not realize that they say the same thing, and so he accepts one while rejecting the other. Since applying strong disquotation gives us the contradictory result that Pierre both believes and does not believe one and the same thing, the strong disquotational principles cannot be accepted.[13]

[12] Kripke (1979).

[13] This is just one of many similar examples in the literature. Another is Nathan Salmon's character Sasha, who learns the words *catsup* and *ketchup* from independent ostensive definitions, in which bottles so-labeled are given to him to season his foods at different times. As a result, Sasha comes to learn what catsup is and what ketchup is. However, since the occasion never presents itself, no one ever tells him that the two words are synonymous, which of course they are. As a result, he does not accept the sentence *Catsup is ketchup*—because he suspects that there may be some, to him indiscernible, difference between them. Nevertheless he understands both words. As Salmon emphasizes, nearly all of us learn one of the words ostensively. The order in which they are learned does not matter, and if either term may be learned ostensively, then someone like Sasha could learn both in that way. But then there will be sentences S_1 and S_2 which differ only in the substitution of one word for the other, which Sasha understands while being disposed to accept only one—just as with Kripke's Pierre. Salmon (1990). See also chapter 15 of Soames (2003).

However, the source of their plausibility should be understood. As I argued in chapter 3 of *Beyond Rigidity*, it is common for an utterance of a sentence to result in the assertion not only of the proposition it **semantically** expresses, but also of other propositions, the contents of which depend on background assumptions in the context. For example, the sentence

2a. Peter Hempel lived on Lake Lane.

might be used in one context to assert the proposition that my former neighbor, Peter Hempel, lived on Lake Lane, while in another it might be used to assert that the famous philosopher, Peter Hempel, lived on Lake Lane. The meaning of the sentence is what is common to what is asserted in **all** normal contexts in which it is used by speakers who understand it. This turns out to be nothing more than the singular, Russellian proposition that is also semantically expressed by (2b).

2b. Carl Hempel lived on Lake Lane.

Since (2a, b) **mean** the same thing, even though speakers who understand them may not realize that they do, anyone who understands both while accepting only one is a threat to principles of strong disquotation. If, in those principles, the proposition p the speaker uses S to express is identified with the proposition **semantically** expressed by S, then the existence of such a speaker falsifies the principles. However, if the principles allow p to be a modestly enriched proposition that the speaker would assert were he to assertively utter S in the context, no counterexample may result.

Thus, small differences in formulation can affect whether or not the principles are compatible with certain problematic examples.[14] When stated in terms of the semantic contents of sentences, strong disquotational principles are straightforwardly false. When stated in terms of descriptively enriched propositions that speakers would use sentences to assert in particular contexts, the principles are more plausible. Unfortunately, these principles are often either left implicit or stated imprecisely, with the resulting danger of equivocation. If Kripke's implicit use of strong disquotation in *Naming and Necessity* is taken as involving a modestly **enriched** proposition that speakers might naturally use the sentence *Hesperus is Phosphorus* to assert—say the proposition that the bright object, Hesperus, seen in the evening is the bright object, Phosphorus, seen in the morning—then his conclusion that **this proposition** is knowable only aposteriori is correct, and the needed version of strong disquotation is **not** subject to immediate falsification. However, this way of taking the argument is of no help to the larger project of vindicating the necessary aposteriori—since the **enriched** propositions speakers associate with (1a) are **not** necessary truths. On the other hand, if we focus on the necessary proposition that the sentence **semantically** expresses, then the strong disquotational principles needed for Kripke's argument cannot be accepted. Either way, when equivocation is avoided, Kripke's use of examples like (1a) to explain the necessary aposteriori fails.[15]

[14] Thanks to Mike McGlone for helping me appreciate this point.

[15] The best one might do, I think, would be to imagine an assertive utterance of *if Hesperus exists and Phosphorus exists, then Hesperus is Phosphorus* in which the speaker asserted the enriched

The Strong Two-Dimensionalist route to P3

A different, more contemporary, route to the problematic principle P3 is provided by a view I call *strong two-dimensionalism*. The prototypical strong two-dimensionalist takes metaphysical possibility to be the only kind of possibility;[16] he takes the function of evidence required for aposteriori knowledge of p to be that of ruling out possible circumstances in which p is false;[17] and he is inclined to analyze propositions as sets of possible world-states (though this last is not strictly required).[18] Given these commitments, he has little choice but to try to explain away the necessary aposteriori as an illusion.

Before getting into this explanation, and the problems with it, it may be worthwhile to spend a few moments distinguishing strong two-dimensionalism from other views to which the adjective *two-dimensionalist* is sometimes attached. First, there is what we might call *benign two-dimensionalism*. Roughly put, this is the view that there are two dimensions of meaning—character and content. The former is a function from contexts of utterance (which include possible world-states in which expressions may be used) to contents. The latter either is, or determines, a function from circumstances of evaluation (again including possible world-states) to extensions. Characters, which are occasionally referred to as *two-dimensional intensions*,[19] are, as David Kaplan has taught us, crucial to the semantics of context-sensitive expressions and the sentences that contain them. It is Kaplan who gave us benign two-dimensionalism, the *locus classicus* of which is his "Demonstratives."[20] In Kaplan's benign sense, "we are all two-dimensionalists now."[21]

proposition expressed by *If the bright object, Hesperus, seen in the evening exists and the bright object, Phosphorus, seen in the morning exists, then the bright object, Hesperus, seen in the evening is the bright object, Phosphorus, seen in the morning.* This proposition is, arguably, both an example of the necessary aposteriori and something which might be predicted to be aposteriori by appropriately formulated strong disquotationalist principles appealing to enriched propositions asserted by speakers. Nevertheless, there appears to be little prospect of finding any formulation of strong disquotationalist principles that both avoids all falsifying counterexamples, and explains the aposteriority of all Kripke-style examples of the necessary aposteriori. Thanks to Ben Caplan for a useful discussion of this point.

[16] See, for example, Chalmers (1996), pp. 136–8, and Jackson (1998), pp. 67–74.

[17] See, David Lewis, "Elusive Knowledge," in Lewis (1999), pp. 422–3.

[18] See Robert Stalnaker, "Assertion," and David Lewis, "Elusive Knowledge." Although Chalmers and Jackson are not as explicit in identifying propositions with sets of metaphysically possible world-states, their views naturally suggest such an identification. For example, see Jackson (1998), pp. 71–2 and 75–77.

[19] Ibid., p. 10 of the introduction to Stalnaker (1999).

[20] Kaplan (1989).

[21] That said, there are some quite misleading passages, as well as some (in my opinion) ill-considered doctrines, in "Demonstratives" that provided fertile ground for the later development of what I call below *ambitious two-dimensionalism.* Examples of misleading passages include those (on pp. 538–9) in which Kaplan suggests that logical truth is a form of apriori truth, and that the bearers of logical truth (and perhaps also of apriori truth) are characters, whereas the bearers of necessity are contents. An example of an ill-considered doctrine is his permissive use of *dthat* as a

In recent years, however, *two-dimensionalism* has come to stand for something more pointed and specific—a cluster of views that build on Kaplan's benign two-dimensionalism, while going beyond it in philosophically significant ways. The defining characteristic of *ambitious two-dimensionalism*, as we might call it, is the attempt to use a Kaplan-like distinction between content and character to explain, or explain away, all instances of the necessary aposteriori and the contingent apriori.[22] The central tenets of this view are the following:

Tenets of ambitious two-dimensionalism

T1. Each sentence is semantically associated with a pair of semantic values—primary intension and secondary intension. The primary intension of S is, in some versions of two-dimensionalism, its Kaplan-style character. In others, it is a proposition which is true with respect to all and only those contexts C to which the character of S assigns a proposition true at C. The secondary intension of (or proposition expressed by) S at a context C is the proposition assigned by the character of S to C.

T2. Understanding S consists in knowing its character (and also knowing which proposition is its primary intension, in those versions of two-dimensionalism in which primary intension is taken to be a proposition true in all and only those contexts to which the character assigns a truth). Although this knowledge, plus complete knowledge of the context C, would give one knowledge of the proposition expressed by S in C, one does not always have complete knowledge of C. Since we never know all there is to know about the designated world-state of C, sometimes we do not know precisely which proposition is expressed by S in C. However, this does not stop us from using S correctly in C.

T3a. Examples of the necessary aposteriori are sentences the secondary intensions of which are necessary, and the characters of which assign false propositions to some contexts. (In versions of two-dimensionalism which identify primary intensions with propositions related to characters, these propositions are contingent.)

T3b. Examples of the contingent apriori are sentences the secondary intensions of which are contingent, and the characters of which assign true propositions to every context. (In versions of two-dimensionalism which identify primary intensions with propositions related to characters, these propositions are necessary.)

vehicle of achieving direct reference (defended on p. 536). The two combine in an unfortunate way in his treatment of the necessary aposteriori and the contingent apriori in Remark 10 of section XIX. In my view, all of this represents a problematic step beyond Kaplan's lasting achievement of benign two-dimensionalism in the direction of the more suspect ambitious two-dimensionalism. (Though Kaplan has never been a full fledged two-dimensionalist in any sense, because he has always rejected all descriptive or indexical analyses of proper names and natural kind terms.) All of these matters, and more, are spelled out in Soames (2005).

[22] For an illuminating early investigation of this strategy for dealing with the necessary aposteriori and the contingent apriori, see Davies and Humberstone (1980).

T4a. All proper names and natural kind terms have their reference semantically fixed by descriptions not containing proper names or natural kind terms.

T4b. These names and natural kind terms are synonymous with context-sensitive, rigidified descriptions (using *dthat* or *actually*).[23]

The core philosophical ideas motivating ambitious two dimensionalists are expressed by T3a and T3b. The necessary aposteriori and the contingent apriori are regarded as posing philosophical problems to which T3a and T3b are thought to provide the answers. Precisely what problems are posed, and what these answers amount to, depend on which ambitious two-dimensionalist view is in question. Roughly speaking, these views come in two main varieties—*strong two-dimensionalism* (very ambitious) and *weak two-dimensionalism* (ambitious, but less so). As indicated earlier, strong two-dimensionalism tends to be driven by three philosophical commitments: (i) the conviction that metaphysical possibility is the only genuine kind of possibility, (ii) the view that the function of evidence required for aposteriori knowledge of a proposition p is that of ruling out possibilities in which p is false, and (iii) the view that propositions are sets of possible world-states. Given (i), plus either (ii) or (iii), one has no choice but hold that no necessary proposition is ever knowable only aposteriori. (In the case of (iii) this is because there is only one necessary proposition, which surely is knowable apriori, while in the case of (ii) it is because there are no possibilities to be ruled out in which necessary propositions are false.) It follows, according to strong two-dimensionalism, that if a sentence S is an instance of the necessary aposteriori, it is **not** because the proposition that S expresses, its so-called secondary intension, is both necessary and knowable only aposteriori. Rather, it is because the secondary intension of S is necessary, whereas its primary intension is contingent. On this view, ***it is a necessary truth that S*** and ***it is knowable only aposteriori that S*** are jointly true, but the proposition said to be necessary is not the one reported to be knowable only aposteriori. In general, when one says, ***Jones knows, or knows apriori, that S***, what one reports is that Jones knows p (or knows p apriori), where p is the primary, rather than the secondary, intension of S. Similarly, when one says ***it is knowable apriori that S*** or ***it is knowable only aposteriori that S***, the proposition one reports on is the primary intension of S, not its secondary intension.[24]

[23] The character of ***dthat [the D]*** is a function from contexts to the denotation o of ***the D*** in the context; propositions expressed by sentences containing ***dthat [the D]*** are singular propositions about o. The character of ***the x: actually Dx*** is a function from contexts C to the property of being the unique object which "is D" in C_w (the world-state of C); propositions expressed by sentences containing the description are singular propositions about C_w.

[24] Typically strong two dimensionalists identify the primary intension of a sentence S not with its character but with a proposition that is true at an arbitrary context iff the character of S assigns that context a secondary intension true at the context. (Since ordinary indexicals like 'I', 'you', today' and so on can cause problems here, it is best, when getting the flavor of this view, to put such indexicals aside—as is done by Davies and Humberstone, for example.)

This view is strongly suggested by well-known works of leading two-dimensionalists like Frank Jackson, David Lewis, and David Chalmers.[25] However, it is not the only form of ambitious two dimensionalism. Weak two dimensionalism, which rejects both (ii) and (iii) above, does not hold, for example, that no necessary proposition is knowable only aposteriori; nor does it construe knowledge and other propositional attitude ascriptions ***x knows/knows apriori/believes etc. that S*** as reporting that the agent bears the relevant attitude to the primary intension of S. Rather, it adopts the more familiar view that these ascriptions report relations between the agent and the secondary intension of, or proposition expressed by, S. Since weak two-dimensionalists recognize that for some sentences S, ***it is knowable only aposteriori that S*** and ***it is a necessary truth that S*** are jointly true, they also recognize that some propositions are both necessary and knowable only aposteriori. What makes them (ambitious) two-dimensionalists is their attempted explanation of this fact. According to weak two-dimensionalism, for all necessary propositions p, p is both necessary and knowable only aposteriori iff (i) p is knowable **in virtue of** one's justifiably accepting some context-sensitive meaning (character) M (and knowing that it expresses a truth), where M is such that (a) it assigns p to one's context, (b) it assigns a false proposition to some other contexts, and (c) one's justification for accepting M (and believing it to express a truth) requires one to possess empirical evidence, and (ii) p is knowable **only** in this way.

Although the differences between strong and weak two-dimensionalism are subtle, they are also far-reaching—too far-reaching to be discussed in this short space.[26] The important point for us is that both of these views are friendly to the principle P3—suggested by Kripke's problematic response to the general objection to the necessary aposteriori discussed in lecture 3 of *Naming and Necessity*.

> P3. When empirical evidence is required (by the agent) for the truth of ***a knows that S***, its function is always to rule out possibilities. However, sometimes the possibilities to be ruled out are **not** those in which the proposition expressed by S is false; instead they are possibilities in which a certain related proposition is false.

The rationale for P3 is particularly clear for the strong two-dimensionalist. According to him, ***a knows that S*** reports the agent's knowledge, not of the secondary intension of S (the proposition S expresses), but of the primary intension of S. Hence,

[25] Robert Stalnaker adopts a similar position in "Assertion." The main difference between Stalnaker and other strong two-dimensionalists is that he presents his version of strong two-dimensionalism—there are sentences that are examples of the necessary aposteriori, but no propositions are both necessary and knowable only aposteriori—in the form of a pragmatic theory about what is asserted by utterances of such sentences, rather than in the form of a two-dimensional semantic theory. For this reason, he is not committed to T1–T4, or to a special semantic analysis of attitude ascriptions. Nevertheless, the pragmatic theory he invokes contains an analog of the standard two-dimensionalist distinction between primary and secondary intension, and he uses the analog of primary intension to attempt to explain what is known aposteriori when Kripke-style instances of the necessary aposteriori are used in conversation.

[26] They are, however, discussed in Soames (2005).

the evidence needed by the agent to support the truth of the knowledge ascription is evidence ruling out possibilities in which the primary intension of S is false. Since, according to the strong two-dimensionalist, this proposition will always be contingent when S is an example of the necessary aposteriori, the necessity of the proposition expressed by S poses no threat to, or puzzle regarding, the requirement that the agent possess empirical evidence in these cases.

Although a similar story supporting P3 can be obtained in the case of weak two-dimensionalism, both the positive story, and the difficulties with it, are less straightforward and more complicated than with strong two dimensionalism. For that reason, I will here focus, in what follows, exclusively on strong two-dimensionalism. Having motivated the strong two-dimensionalist's acceptance of P3, I will bring out the fundamental difficulties that make the position untenable. (In stating and criticizing strong two-dimensionalism I leave ordinary indexicals like 'I', 'you', 'he' and 'now' aside, since these raise additional, independent problems and complications.)[27]

In the interest of explicitness, I fill out the sketch of strong two-dimensionalism by adding theses T5 and T6 to T1–T4.

T5. ***It is a necessary truth that S*** is true with respect to a context C iff the secondary intension of S in C is true with respect to all world-states that are possible relative to C. By contrast, ***it is knowable apriori that S*** is true with respect to C iff in C, the primary intension of S is knowable apriori; ***x knows/believes that S*** is true of an individual i in C iff in C, i knows/believes the primary intension of S. Similarly for other modal and epistemic operators.

T6. S is an example of the necessary aposteriori iff the secondary intension of S (with respect to C) is a necessary truth, but the primary intension of S, though knowable, is not knowable apriori. In all such cases, the primary intension of S is contingent—that is, there are contexts C* to which the character of S assigns a proposition that is false in C*. Thus, examples of the necessary aposteriori express necessary truths in our actual context, while expressing falsehoods in other contexts. Primary intensions of these sentences are not knowable apriori because we require empirical information to determine that our context is not one to which the character assigns a falsehood.

These points are illustrated by the sentences in (3).

3a. The actual husband of Stephanie Lewis was the actual author of *Counterfactuals*.
b. The husband of Stephanie Lewis was the author of *Counterfactuals*.

The two rigidified descriptions in (3a) rigidly designate David Lewis. Hence, the **secondary intension** of (3a) is a necessary truth.[28] By contrast, the proposition

[27] The nature of, and intractable difficulties with, all major forms of ambitious two-dimensionalism are discussed in Soames (2005).

[28] To keep things simple I will ignore world-states in which David does not exist. I will also, in discussing this particular example, make the simplifying assumption that the names occurring in

expressed by (3b) is contingent, and obviously knowable only aposteriori. Since (3b) expresses the same proposition in every context of utterance, this proposition—the secondary intension of (3b)—is taken to be its primary intension as well. Now note that (3a) expresses a truth in all and only those contexts in which (3b) expresses a truth. This means that the **primary intension** of (3a) is necessarily equivalent to the contingent, aposteriori proposition that is both the primary and secondary intension of (3b). They may even be identified, since, for the prototypical strong two-dimensionalist, propositions are sets of possible world-states. However, even if one were to resist this identification, one would have to acknowledge the trivial equivalence of these propositions. Anyone who understands both sentences knows that they have the same truth value in any context in which they are used, and anyone who apprehends both the primary intension of (3a) and the primary/secondary intension of (3b) can see immediately that they are equivalent. It follows that since the latter is knowable only aposteriori, the former is so as well. As a result, the strong two-dimensionalist maintains that sentence (3a) is an example of the necessary aposteriori, even though it is not associated with any one proposition that is both necessary and knowable only aposteriori.

This illustrates one of the central theses of strong two dimensionalism: **no single proposition can be both necessary and knowable only aposteriori**. The thought that there are such propositions is due to an equivocation. When S embeds under a modal operator, its secondary intension is relevant; when S embeds under an epistemic operator, its primary intension is relevant. Since names and natural kind terms are analyzed as rigidified descriptions, the two intensions will be different whenever S contains any of these expressions. Hence, the strong two-dimensionalist believes he can explain away all Kripkean examples of the necessary aposteriori on the model of (3a).

However, he is wrong about this. The pattern of explanation offered suffers from fatal flaws, as is shown by the following four arguments against strong two-dimensionalism.[29]

Argument 1

1. According to strong two-dimensionalism, epistemic attitude ascriptions ***a V's that S*** report that the agent bears the relation expressed by V to the primary intension

the sentence are non-indexical expressions with constant characters. Although this assumption runs contrary to two-dimensionalist doctrine, taking it for granted here will allow us to focus on the difference between rigidified and unrigidified descriptions in the two-dimensionalist explanation of the necessary aposteriori. Later, in criticizing the two-dimensionalist framework, I will remove the simplifying assumption.

[29] In stating arguments 1 and 2, I will take for granted the prototypical strong two-dimensionalist identification of propositions with sets of possible world-states. However, this assumption could, in principle, be relaxed without affecting the final conclusions of these arguments. All one needs is observations of the following sort: that for any possible agent a and context C in which a finds himself, a will accept the character of ***Actually S***, and believe it to express a truth, only if a accepts the character of S, and believes it to express a truth. Using this, one can reach the conclusions of arguments 1 and 2 without appealing to the premise that necessarily equivalent propositions are identical.

of S—that is, to the proposition that, in effect, says of the character of S that it expresses a truth.

2. Since for every context C, the character of (3a) expresses a truth with respect to C iff the character of (3b) does too, the two primary intensions are identical, and the ascriptions ***a V's that the actual husband of Stephanie Lewis was the actual author of Counterfactuals*** and ***a V's that the husband of Stephanie Lewis was the author of Counterfactuals*** are necessarily equivalent.
3. Hence, the truth value of *Necessarily [if the actual husband of Stephanie Lewis was the actual author of Counterfactuals and Mary believes that the actual husband of Stephanie Lewis was the actual author of Counterfactuals, then Mary believes something true]* is the same as the truth value of *Necessarily [if the actual husband of Stephanie Lewis was the actual author of Counterfactuals and Mary believes that the husband of Stephanie Lewis was the author of Counterfactuals, then Mary believes something true]*. Since the latter modal sentence is false, so is the former.
4. Similarly, the truth value of *Necessarily [if Mary believes that the actual husband of Stephanie Lewis was the actual author of Counterfactuals, and if that belief is true, then the actual husband of Stephanie Lewis was the actual author of Counterfactuals]* is the same as the truth value of *Necessarily [if Mary believes that the husband of Stephanie Lewis was the author of Counterfactuals, and if that belief is true, then the actual husband of Stephanie Lewis was the actual author of Counterfactuals]*. Since the latter modal sentence is false, so is the former.
5. Since, in fact, the initial modal sentences in steps 3 and 4 are true, strong two-dimensionalism is false. Lesson: Don't assign modal and epistemic operators different objects, since if you do, you won't assign the correct truth conditions to sentences in which they interact.

Argument 2

1. According to strong two-dimensionalism, epistemic attitude ascriptions ***a V's that S*** report that the agent bears the relation expressed by V to the primary intension of S—that is, to the proposition that, in effect, says of the character of S that it expresses a truth.
2. According to strong two-dimensionalism, names are synonymous with rigidified descriptions. Let o be an object uniquely denoted by the nonrigid description ***the D***, let n be a name of o, and let the strong two-dimensionalist analysis of n be ***the actual D***. Suppose further that ***John believes that n is D*** is true.
3. Let w be a world-state in which some object other than o is uniquely denoted by ***the D***, and in which John does not believe of o that it "is D," though he does believe the proposition expressed by ***The D is D***.
4. According to strong two-dimensionalism the truth values of (a) and (b) must be the same.
   a. ***Although John truly believes that n is D, had the world been in state w, n would not have been D and John would not have believed that n was D.***

b. ***Although John truly believes that the actual D is D, had the world been in state w, the actual D would not have been D and John would not have believed that the actual D was D.***

5. Since, according to strong two-dimensionalism, ***John believes that the actual D is D*** and ***John believes that the D is D*** are necessarily equivalent, occurrences of the latter can be substituted for occurrences of the former in (b) without changing truth value. Hence, if (a) and (b) are true, then (c) must also be true.

   c. ***Although John truly believes that the D is D, had the world been in state w, the actual D would not have been D and John would not have believed that the D was D.***

6. In fact, however, (a) is true and (c) is false. Hence strong two-dimensionalism is false. Lesson: Analyzing names as rigidified descriptions only compounds the problem revealed in argument 1.

Argument 3

1. In point of fact, (a) entails (b).

   a. ***John truly believes that n is D, but had the world been in state w, n would not have been D and John would not have believed that n was D.***

   b. ***There is an x such that John truly believes that x is D, but had the world been in state w, x would not have been D and John would not have believed that x was D.***

2. If the semantics of strong two-dimensionalism were correct, there would be no such entailment—since (b) could be false when (a) was true. (There is no distinction between primary and secondary intensions for variables, though strong two-dimensionalists insist that there is such a distinction for names.)

3. So, the semantics of strong two dimensionalism is incorrect. Lesson: Strong two-dimensionalism misses the following semantic fact: if ***John believes that n is F*** is true at world-state w at which n designates o, then at w John believes of o that it "is F", and ***John believes that x is F*** is true at w with respect to an assignment of o to 'x'.

Argument 4

1. Let S be an example of the necessary aposteriori that the strong two-dimensionalist characterizes as such. Let it further be the case that ***John does not know that S*** is true because John lacks the empirical information required for such knowledge.

2. Then, according to the strong two-dimensionalist, (a) is true.

   a. ***It is a necessary truth that S but it is not knowable apriori that S, and although it is knowable that S, John does not know that S.***

3. In point of fact, (a) entails (b).

   b. *There is some necessary truth p, which is not knowable apriori, and although p is knowable, John does not know p.*

4. (b) contradicts the strong two-dimensionalist's central thesis that no single proposition is both necessary and knowable only aposteriori. In addition, it conflicts with his identification of propositions with sets of possible world-states, since if p is the unique necessary truth, then John surely knows it.
5. Since the strong two-dimensionalist accepts the truth of (a), he must declare that, according to his semantic theory, (b) is not a consequence of (a).
6. Since (b) clearly is a consequence of (a), the strong two-dimensionalist's semantic theory is incorrect. Lesson: Objectual variables ranging over propositions can be objects of modal predicates and propositional attitude verbs. Since these are not associated with distinct primary and secondary intensions, they cannot be given a two-dimensionalist treatment.

Other arguments that make crucial use of ordinary indexicals, like *I*, *you*, *he*, and *now* could also be given against strong two-dimensionalism.[30] But there is no need to go in for further criticism. Strong two-dimensionalism—as a semantic theory of names, natural kind terms, and modal and epistemic predicates and operators—is false, as is the account it gives of the necessary aposteriori. As I have indicated, there are, of course, other forms of ambitious two-dimensionalism that warrant investigation in their own right. Since I cannot undertake that task here, I will simply report that they have sorrows of their own.[31] In addition, it is worth noting that, quite independent of its particular defects, weak two-dimensionalism abandons the idea of treating the necessary aposteriori as an illusion, in favor of trying to give a benign explanation of how it is that a single proposition can be both necessary, and knowable only aposteriori. But once this step is taken, there is really no need for a special two-dimensional semantic treatment of names and natural kind terms, since a plausible Kripkean explanation of the necessary aposteriori is readily available. To understand this explanation, we must return to the initial misstep that started us down the blind allies of strong disquotationalism and strong two-dimensionalism.

The Proper Response to the Skeptical Objection to the Necessary Aposteriori

In responding to the skeptical objection to the necessary aposteriori, Kripke had the choice of rejecting either the skeptic's premise P1 or his premise P2. Although his discussion of this point in *Naming and Necessity* is not completely transparent, we saw that the crucial passages encouraged a reading in which P1 was rejected and replaced by P3 and P4. Since indicating this, I have been trying to illustrate the futility of P3 by cataloging the problems with different theoretical strategies for implementing it. Having indulged in this negativity, I now turn to the positive solution that has been staring us in the face all along.

As I indicated earlier, Kripke's views about the necessary aposteriori are connected to his views about essential properties. He argues that we know **apriori** that various

[30] These are presented in Soames (2005). [31] Soames (2005).

properties and relations are essential to anything that has them. This means that certain propositions which predicate these properties and relations of objects are such that we know apriori that **if** they are true, **then** they are necessarily true. Still, finding out whether they are true requires empirical investigation. According to this way of looking at things, in order to find out whether certain things are true with respect to **all** possible states of the world, and other things are true with respect to **no** possible states of the world, we sometimes must **first** find out what is true with respect to the **actual** state of the world. Sometimes in order to find out what could or could not be, we first must find out what is. This will seem problematic only if one has restricted the ways things could coherently be **conceived** to be to ways things **really** could be—that is only if one has restricted epistemic possibility to metaphysical possibility. Although the passages in lecture 3 of *Naming and Necessity* that we have been discussing may seem to show Kripke backsliding on this point, they do not, in my opinion, negate the central lesson of his work that one must sharply distinguish these two kinds of possibility. Thus, the proper response to the Kripkean objector is to reject the skeptic's premise P2.[32]

For Kripke, what is epistemically possible is not always metaphysically possible. Here, it is helpful to remember that, for him, possible states of the world are not alternate concrete universes, but abstract objects—maximally complete ways the real concrete universe could have been. They are, in effect, maximally complete properties that the universe could have instantiated. Thinking of them in this way suggests an obvious generalization. Just as there are properties that certain objects could possibly have had and other properties they could not possibly have had, so there are certain maximally complete properties that the universe could have had—possible states of the world—and other maximally complete properties that the universe could not have had—impossible states of the world. Just as some of the properties that objects could not have had are properties that one can coherently conceive them as having, and that one cannot know apriori that they do not have, so some maximally complete properties that the universe could not have had (some metaphysically impossible states of the world) are properties that one can coherently conceive it as having, and that one cannot know apriori that it does not have. Given this, one can explain the informativeness of certain necessary truths as resulting from the fact that learning them allows one to rule out certain impossible, but nevertheless

[32] Kripke's footnote 72, toward the end of the main passage under discussion, shows that even there he was aware of the importance of the distinction between epistemic and metaphysical possibility. He says, referring to some of the remarks we have been discussing: "Some of the statements I myself make above may be loose and inaccurate in this sense. If I say, 'Gold *might* turn out not to be an element,' I speak correctly; 'might' here is *epistemic* and expresses the fact that the evidence does not justify *a priori* (Cartesian) certainty that gold is an element. I am also strictly correct when I say that the elementhood of gold was discovered *a posteriori*. If I say, 'Gold *might have* turned out not to be an element,' I seem to mean this metaphysically and my statement is subject to the correction noted in the text." Here, it is important to remember that the footnotes were added to the lectures by Kripke after they were given, and a written transcript had been produced. I believe that when writing the footnote he noticed that his discussion had neglected the distinction between epistemic and metaphysical necessity, and he wished—without changing the text—to call attention to his commitment to it.

conceivable, states of the world. Moreover, one can explain the function played by empirical evidence in providing the justification needed for knowledge of necessary aposteriori truths. Empirical evidence is required to rule out certain impossible, but nevertheless coherently conceivable and epistemologically relevant, world-states which (i) cannot be known apriori not to obtain, and (ii) are such that the necessary aposteriori truths are false with respect to those world-states.[33] Thus, by expanding the range of epistemically conceivable states of the world to include some that are metaphysically impossible, one can accommodate Kripkean examples of the necessary aposteriori. This, not two-dimensionalism, is the true lesson of his seminal discussion of this important category of truths.[34]

References

Chalmers, David (1996). *The Conscious Mind* (New York and Oxford: Oxford University Press).

Davies, Martin and Humberstone, Lloyd (1980). "Two Notions of Necessity," *Philosophical Studies*, 38.

Jackson, Frank (1998). *From Metaphysics to Ethics* (Oxford: Oxford University Press).

Kaplan, David (1989). "Demonstratives," in Almog, Perry, and Wettstein (eds.), *Themes From Kaplan* (New York and Oxford: Oxford University Press).

Kripke, Saul (1979). "A Puzzle about Belief," in A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel).

____(1980). *Naming and Necessity* (Cambridge MA: Harvard University Press).

Lewis, David (1996). "Elusive Knowledge," *The Australasian Journal of Philosophy* 74, 1996; reprinted in Lewis (1999).

____(1999). *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).

[33] I here assume that names (unlike definite descriptions such as *the stuff out of which this table, if it exists, is constituted*) rigidly designate the same thing with respect to all world-states, metaphysically possible or not.

[34] There are, of course, other worries—independent of the characterization of "worlds" as abstract or concrete—that some philosophers have raised about recognizing epistemologically possible, but metaphysically impossible, world-states. For example, on pp. 136–8 of Chalmers (1996), he worries about the view there are world-states that are "entirely conceivable" but not metaphysically possible at all. He argues that such a view would render metaphysical possibility arbitrary and mysterious, by divorcing it from conceivability, and also that if there were epistemically possible world-states that were not metaphysically possible, "we could never know it." However, these arguments are mistaken, as I undertake to show in Soames (2005). Another philosopher who rejects the epistemologically possible, over and above the metaphysically possible, is Robert Stalnaker. His main concern seems to be that of grounding all intentionality (including all forms of belief and knowledge) on non-intentional modal facts involving metaphysical possibility. Since, as he recognizes, admitting epistemic possibility would spoil this project, he declines to do so. (See in particular pp. 24–5 of chapter 1 of Stalnaker (1984).) As I see it, however, this concern is undermined by the fact that his attempted explication of the intentional in terms of the metaphysically possible fails for independent reasons (see, for instance, Speaks (2003)), and also by the fact that his pragmatic two-dimensionalist model of discourse, which is restricted to metaphysically possible world-states, cannot account for straightforward conversational uses of the necessary aposteriori (as I argue in Soames (2005)).

Salmon, Nathan (1990). "A Millian Heir Rejects the Wages of *Sinn*," in C. A. Anderson and J. Owens, eds., *Propositional Attitudes: The Role of Content in Logic, Language, and Mind* (Stanford, CA.: CSLI).

Soames, Scott (2002). *Beyond Rigidity* (New York: Oxford University Press).

____(2003). *Philosophical Analysis in the Twentieth Century, Vol. 2: The Age of Meaning* (Princeton, NJ: Princeton University Press).

____(2005). *Reference and Description: The Two-Dimensionalist Attempt to Revive Descriptivism* (Princeton, NJ: Princeton University Press)

Speaks, Jeff (2003). *Three Views of Language and the Mind* (unpublished Princeton dissertation, 2003).

Stalnaker, Robert (1999). *Context and Content* (Oxford: Oxford University Press).

____(1984). *Inquiry* (Cambridge, MA: MIT Press).

13

Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics

Robert Stalnaker

Beginning more than twenty-five years ago, two-dimensional modal semantics has been applied to the interpretation of speech and thought in a number of different ways. More recently, the two-dimensional semantic apparatus has been deployed (by Frank Jackson and David Chalmers, among others) in philosophical arguments about the role of conceptual analysis and reductive explanation, and in the clarification of the notions of a priori knowledge and truth. Different philosophers have applied the apparatus to examples in the same way, but have given contrasting interpretations of those applications, and have drawn different philosophical conclusions about their significance. My intention in this paper is to try to clarify the question of interpretation, and the contrasting ways of answering it. I will defend one kind of interpretation, but my main aim is to draw a contrast between two very different ways of thinking about intentionality that I think are implicit in the different ways of understanding the framework. I will begin with a look back at my own early attempts to deploy this framework, at the way I understood the problem to which it was a response, and the way I was interpreting it. Second, I will sketch a contrasting interpretation, which generalizes David Kaplan's semantics for context-dependent expressions. Third, I will look at David Chalmers's different way of contrasting the different interpretations, and at the account that he defends. Finally, I will describe and criticize the internalist approach to intentionality that I think is required by the alternative interpretations of the two-dimensional semantic framework.

The occasion for this paper was the conference on two-dimensional semantics at the ANU in February, 2002, but it was written after the conference. The paper that was the basis of both the talk I gave there and at the Barcelona conference on two-dimensional semantics the previous June was Stalnaker (2004b), which discusses some similar themes. I am grateful to the participants in both of these conference for the high level of discussion that helped me to get clearer about the different ways of understanding and using the two-dimensional semantic framework. Thanks particularly to Philip Pettit for his comments on my paper at the ANU conference.

1. The "Assertion" Story

In my paper "Assertion,"[1] I began with a simple abstract but intuitive picture of what it is to say or think something. According to this picture, a representation is a way of distinguishing between possibilities. As Frank Jackson puts it, "to represent is to make a division into what accords with, and what does not accord with, how things are being represented as being."[2] A *proposition*—the content of a representation—can be modeled, according to this picture, by the set of possible situations that *are* the way the world is being said to be. These possible situations are the *truth conditions* of the representation—the conditions that would have to obtain for the proposition to be true. An *assertion* can be understood as a proposal to exclude from the possible situations compatible with the context those in which the proposition asserted is false.

Two-dimensional semantics came into the picture as a response to a problem that this conception of representation brought into focus. The problem is this: with some statements, there is a tension between global intuitions about the information that the statement conveys, as represented by the possibilities that the statement seems to exclude, and what semantic theories that are otherwise well motivated say about the truth conditions of the statement. The tension is most acute with statements that seem to be informative (and so to exclude possibilities), but also necessarily true (and so to exclude no possibilities). The clearest cases of this kind are the necessary a posteriori statements that Saul Kripke brought to our attention in *Naming and Necessity*. It seems intuitively clear, for example, that identity statements with proper names such as "Hesperus is Phosphorus" and statements about the nature of natural substances such as "water is a compound of hydrogen and oxygen" convey substantive information about the world, but it also seems that such statements say something that could not possibly be false.

The first step to get clear about the problem is to ask what information it is that statements of this kind seem to be conveying—what kinds of possible situations the statement seems to be excluding. If Daniels has reason to tell O'Leary that Hesperus is Phosphorus, it must be that he thinks that O'Leary doesn't know it already—that certain possibilities need to be excluded, and that saying that Hesperus is Phosphorus will succeed in excluding them. If O'Leary thinks that Hesperus might not be Phosphorus, what does he think the world might be like? If we can give a plausible answer to this question, the second step is to ask how it is that the statement "Hesperus is Phosphorus," which our semantic theory tells us is a necessary truth, manages to do the job of excluding those possibilities.

If O'Leary doesn't know that Hesperus is Phosphorus, then it seems reasonable to think that possible worlds that satisfy the following description are compatible with his knowledge: There is a heavenly body that appears in the evening where Venus in fact appears, and a distinct heavenly body that appears in the morning where in fact Venus appears. The first has come to be called "Hesperus" and the second has come

[1] Stalnaker (1978). [2] Jackson (2001: 617).

to be called "Phosphorus," so that what people would be saying in this kind of world if they were to say "Hesperus is Phosphorus" would be false.

The answer to this first question is much the same as the answer that a Fregean might give. The Fregean might describe such a possible world by saying that it is one in which distinct objects are presented by two of the modes of presentation which in the actual world present the same thing—Venus. On the Fregean view, the thought expressed, in the actual world, by the statement "Hesperus is Phosphorus" is a contingent proposition that is false in the possible world described, and it is the very same proposition as the one that would be expressed by someone in that counterfactual possible world who said "Hesperus is Phosphorus" there. But my aim was to reconcile the fact that the statement is informative with a direct reference account of the semantics for names according to which the possible world we have described is not one in which Hesperus is distinct from Phosphorus, since there is no such world. On this account of the semantics of names, the possible world we have described is one that differs from the actual world not only in its astronomical facts, but also in its semantic facts: it is a world in which the expression "Hesperus is Phosphorus" expresses a different proposition. The direct reference theory of names gave an externalist account of the facts that determine reference: statements containing names have the content that they have because of the way speakers using them are causally connected with things in the world. Consequently, in possible worlds where the astronomical facts are different, the semantic values of names referring to astronomical bodies may be different, and so different propositions may be expressed with those names.

Externalist accounts of names, and more generally of propositional content, focused attention on the fact that it is a matter of contingent fact that the words we use have the meaning and content that they have. This is of course not a fact that is restricted to externalist theories—any theory of speech and thought must give an account of the facts in virtue of which marks, sound patterns, and states of the brains of people and animals have the representational properties that they have. But externalist theories of intentionality make this fact particularly salient. The two-dimensional framework was deployed, in the first instance, as a piece of descriptive apparatus for representing the way that semantic values depend on the facts. We need two dimensions since we start with the fact that the truth value of a proposition (at least a contingent proposition) depends on the facts. But since the identity of the proposition expressed in a given utterance also depends on the facts, the truth value of the utterance will depend on the facts in two different ways: first, the facts determine what is said; second, the facts determine whether what is said is true. We can represent the two different roles of the facts in determining a truth value with what I called a *propositional concept*: a function from possible worlds to propositions, where a proposition is a function from possible worlds to truth values, or equivalently, a function from a pair of possible worlds to a truth value.

This descriptive apparatus is of interest independently of our particular problem since it is apt for representing the interaction of speakers and addressees in a conversation. The information that speakers take for granted when they speak includes a mix of semantic information and information about the subject matter of a conversation.

Speakers make assumptions about what their addressees know or believe about what they are talking about as well as about what their words mean and what the relevant contextual parameters are relative to which their words will be interpreted. These assumptions will influence what they choose to say, and how they choose to say it. Since the speaker will normally presuppose that he is speaking (and that the addressee knows this), the possible worlds compatible with what is presupposed will be possible worlds in which the utterance event in question takes place, and in which it has a meaning and a content that may be different from the meaning and content that the utterance has in the actual world. In a case where the addressee knows that the utterance event has taken place, but mishears or misinterprets it, the content of what is said, and sometimes even the words that are uttered, will be different in possible worlds compatible with what the addressee believes than they are in the actual world. Cases where someone is ignorant or mistaken about the content of an utterance will be cases where the propositional concept for the utterance, relative to the relevant possible worlds, will be a variable one: the function will determine different propositions relative to different possible worlds. An example: The policeman has stopped a driver, and after examining his driver's license, says, "your license says you need corrective lenses, but you're not wearing your glasses." The driver responds, "I've got contacts." The policeman replies, "I don't care who you know, you have to wear your glasses." The proposition expressed in the actual world by the driver's utterance "I've got contacts" is different from the proposition expressed by that utterance in the possible worlds compatible with the policeman's beliefs, and this can be represented with a variable propositional concept, defined on the relevant possible worlds.

The descriptive apparatus is relevant to our problem since the possible worlds that the problematic statement, "Hesperus is Phosphorus," seems to be excluding are worlds that differ from the actual world both in astronomical and in semantic facts, and so it seemed that a two-dimensional representation of this kind might help us to answer the second question about such problematic statements: how is it that a necessarily true statement could be used to convey contingent information? To approach this question, we consider an intuitively natural context for our example in which Daniels tells O'Leary that Hesperus is Phosphorus, and construct a propositional concept for it, on the assumption that the semantics for names is the way Kripke argued that it is: names like "Hesperus" and "Phosphorus" are rigid designators—they denote the same thing in all possible worlds, and so the proposition actually expressed by "Hesperus is Phosphorus" is necessarily true. But if the facts that determine the reference of the names had been different—for example if the astronomical facts had been as O'Leary thinks they might in fact be, then an utterance of "Hesperus is Phosphorus" would instead have expressed the necessarily false proposition, since the two names would have denoted different things. So if *i* is the actual world, and *j* is the world described above that is the way O'Leary thinks the world might be, then the propositional concept for the utterance, relative to these two possible worlds, will be the one pictured in this matrix:

	i	*j*
i	T	T
j	F	F

Daniels's purpose in stating that Hesperus is Phosphorus is clearly to exclude worlds like *j*—to inform O'Leary that the actual world is not like that, while including worlds like *i*. Neither the necessarily true proposition expressed in the actual world nor the necessarily false proposition expressed in world *j* (according to the semantics we are assuming) accomplishes this, but the *diagonal* proposition—the one that for each world *x* is true in world *x* if and only if the proposition expressed in *x* is true in *x*, seems to be the proposition that does the right job.

The proposal made in "Assertion" was that in special cases, where there was a prima facie violation of certain conversational rules, utterances should be reinterpreted to express the diagonal proposition, rather than the proposition expressed according to the standard semantic rules. The proposal followed the pattern of interpretive reasoning that Paul Grice spelled out in his theory of conversation: Certain maxims of conversation are argued to be truisms required by the general purposes of rational discourse, and so to be common ground among the participants in a conversation. The presumption that such rules are being followed constrains the interpretation of what is said. It is presumed that hearers will try to find a way of understanding what is said that conforms to the maxims, and speakers may exploit this presumption by saying things that would be manifest violations of conversational rules if they were interpreted in a standard way, and so that will require reinterpretation. In the case of our problematic statements, the relevant maxim is that speakers presume that their addressees understand what they are saying. In terms of the two-dimensional apparatus, this presumption will be satisfied if and only if the propositional concept for the utterance is constant, relative to the possible worlds that are compatible with the context. Our problematic example, and all cases of necessary truths that would be informative (in the sense that the addressee does not already know that they are true) will be prima facie violations of this maxim, and so will require reinterpretation. Reinterpreting by taking the diagonal proposition to be the one the speaker intends to communicate brings the statement into conformity with the rule, and seems to give the intuitively correct result.

On this kind of account, diagonal propositions (corresponding to what Frank Jackson calls A-intensions) are derivative from the standard semantics, as it is in the actual world, and in the relevant alternative possible worlds. My project began from the assumption that the standard semantics (which determines what Jackson calls C-intensions) was essentially right—the project was to reconcile it with the fact that statements that were necessary, according to that semantics, could be used to communicate contingent information. Diagonalization was *re*interpretation—interpretation that was parasitic on the standard interpretation, which was assumed to give the right result for what was expressed and communicated in the normal case. If the standard semantics for names did not give the right result in the normal case—that is, if it were not right to say that normally the beliefs one expresses and the information one conveys when one uses proper names are singular propositions about the individual named—then that semantic account would not be defensible. The plausibility of this assumption turns on one's account of intentionality—of what makes it the case that our mental states have the representational properties that they have. It is disagreements about how the problem

of intentionality is to be solved that lie behind disputes about the semantics for names, and I will suggest that this is also the central issue that divides different interpreters of the two-dimensional semantic apparatus.

In several of the early papers in which I applied the diagonalization strategy,[3] I contrasted the use I was making of the two-dimensional apparatus with the use that David Kaplan made of it in his theory of indexicals and demonstratives.[4] Kaplan's theory is a descriptive semantics for a formal language containing context-dependent expressions such as personal and deictic pronouns, tenses, temporal and locative adverbs. In Kaplan's semantics, the meaning of a sentence type (which he called its *character*) is a function from context to content, where content (what is said) is the proposition expressed. So, for example, the meaning of the sentence "I am flying to Canberra tomorrow" is a function that takes a context of utterance in which **a** is the speaker and **t** is the time of utterance into the proposition that is true in possible worlds in which **a** flies to Canberra on the day following time **t**. The thought expressed in a use of that sentence is the proposition determined; the role of the context is to contribute to the means used to express it.

Kaplan's characters are abstract objects that are similar to propositional concepts, but they play a very different role in the explanation of speech, and were not designed to solve the problem to which the diagonalization strategy was applied. In Kaplan's semantics, the paradigm examples of sentences that express necessary a posteriori truths such as "Hesperus is Phosphorus" are not context-dependent, and so have constant character. If an analogue of the diagonal proposition were defined in terms of the character of such a sentence, it would be the same as the content expressed, a necessary truth, and so would not help to explain how such statements can convey contingent information. Kaplan's semantics will play a role in some applications of the descriptive apparatus, since features of context on which content depends are sometimes among the features that hearers may be ignorant or mistaken about, and the reinterpretation strategy may be required in such cases. The two theories are not competing theories for explaining the same phenomena, or competing interpretations of the abstract framework, but complementary theories that use formally similar tools to answer different questions.

One important difference between the two theories is the contrasting roles of the two-dimensional intensions (character, in Kaplan's semantics, propositional concepts in the assertion theory) in the explanation for the fact that an utterance has the content that it has. Suppose we ask why a certain utterance of the sentence "I am flying to Canberra tomorrow" expresses the proposition that is true if and only if RS flies to Canberra on February 19, 2002. The answer is, because the sentence has the character stated above, and the utterance in question was produced by RS on February 18, 2002. One might go on to ask the further question, what made that utterance an utterance of a sentence with that character, but even if there are further questions, the answer we gave is correct. Character precedes content in the order of explanation of the fact that the utterance has the content that it has. But the order is the reverse

[3] Stalnaker (1981b and 1987). [4] Kaplan (1989a).

in the case of the explanation of why an utterance conveys the information that a diagonal proposition represents. Why does "Hesperus is Phosphorus" (uttered in a particular context) convey the contingent information that the heavenly body that appears in the evening and is called "Hesperus" is distinct from the one that appears in the morning and is called "Phosphorus"? The answer begins with the fact that in a world of this kind that is compatible with the context, the semantics, as it is in that world, implies that the sentence expresses a necessary truth, whereas the semantic accounts that hold in other worlds compatible with the context imply that it expresses a necessary falsehood. We explain why the utterance determines the propositional concept that it determines in terms of the content that it has, or would normally have, according to the semantics of the relevant alternative possible worlds. Content (in the various alternative worlds) precedes propositional concept in the order of explanation. The second part of the explanation invokes reinterpretation by diagonalization, but since the diagonal proposition is determined by the propositional concept, the main work of explaining why the utterance conveys the particular content that it conveys is done when we have explained why the utterance determines the propositional concept that it determines. Again, we can ask a further question: what facts about these possible worlds make it the case that the semantics of those worlds determine that the utterance in question expresses the necessarily true, or necessarily false proposition, but even if there are further questions, the answer we gave gives a correct explanation, assuming we are right about the semantics for the sentence in the different possible worlds.

Just to see how the two theories can interact, involving both kinds of explanation, consider an example that involves both diagonalization and demonstratives: Pierre, in London, says "Londres est jolie, mais cette ville-ci n'est pas jolie." (London is pretty, but *this* city is not pretty.) Jacques responds, "Mais cette ville-ci *est* Londres." (But this city *is* London). Jacques's statement communicates to Pierre the contingent information that the world is not the way he thinks it is, not a world in which the city he calls "Londres" is distinct from the city he is currently in. Why does the utterance of this sentence convey this information? Because the semantics implies that this utterance token expresses a necessary truth in worlds in which the place of utterance is London and a necessary falsehood in worlds in which the place of utterance is a city different from London. Since worlds of both kinds are compatible with the context required to interpret Jacques's utterance, the utterance is reinterpreted to express the diagonal proposition, which is true in worlds of the first kind, and false in worlds of the second kind. But why does the semantics imply that this utterance (on the standard interpretation) expresses a necessary truth in worlds of the first kind and a necessary falsehood in worlds of the second kind? Because the semantics (which is common knowledge in the context, and so applies in all the relevant possible worlds) says that "cette ville-ci" is a rigid designator for the city which is the place of utterance, and that "Londres" is a rigid designator for London. So we explain the propositional concept determined by an utterance in terms of the content expressed by that utterance in different possible worlds and the content expressed by the utterance in the different possible worlds in terms of the character that the

sentence used to make the utterance has, and the context in which it is uttered, in those different possible worlds.

2. The Generalized Kaplan Interpretation

Although I have been arguing that the Kaplan semantics and the assertion theory are complementary theories—formally similar in certain respects, but doing quite different jobs—the two theories have often been taken to be slightly different variations on the same theme. Some have proposed that even though the Kaplan semantics as it is does not apply to the phenomena of informative necessary truth, it can be modified and extended so that it does.[5] This kind of project suggests an alternative interpretation of the two-dimensional framework, as applied to our problem.

The idea is to take a Kaplanian character to be a kind of narrow content, for thought as well as for speech.[6] A semantics that fits what I have called the *generalized Kaplan paradigm* treats a much wider range of expressions as context-dependent: almost all descriptive expressions of the language will have a variable character. While in the original Kaplan theory, it was the content determined that was the thought expressed in the use of an expression, in the generalized theory, it is the character (or the A-intension, or diagonal, that it determines) that is the cognitive value of what is expressed. When a person thinks or says that Socrates lived in Athens, or that there is water on Mars, the thought that he has or expresses is a descriptive proposition about whatever the person and city, or substance and planet, are that fit certain descriptions, or that present themselves to the thinker in certain ways. The C-intension determined will be a singular proposition about Socrates and Athens, or a proposition about the actual substance water and the planet Mars, but these are propositions to which the speaker or thinker has only indirect access. The rigidity of the proper names and natural kind terms is the result of a kind of generalized scope device. The character, or two-dimensional intension, for a thought or utterance corresponds to a non-rigid description of a proposition (the C-intension). The A-intension is the proposition that the C-intension that fits this description is true. The content of the thinker's

[5] It is mainly David Chalmers and Frank Jackson that I have in mind as proponents of the generalized Kaplan interpretation, though my rough sketch of the view may not exactly match the way either of them would state it. See Chalmers (1996) and (2002) and Jackson (1998).

[6] Frank Jackson and David Chalmers give different answers to the question whether the two-dimensional apparatus applies to thought as well as to language. Chalmers assumes that mental states as well as utterances are associated with the two different kinds of intension, while Jackson makes the A-intension/C-intension distinction only for linguistic expressions. But as I understand him, Jackson would say that when one makes an assertion, or when one attributes a belief, it is in general the A-intension of the sentence used to make the assertion, or of the sentential clause used to attribute the belief, that is the proposition that the speaker expresses, or that the subject of the belief attribution is said to believe. This is the sense in which, for Jackson, the A-intension represents the cognitive value of an expression. Since for Chalmers, thoughts (mental analogs of sentences) themselves have the two kinds of intension, it is slightly less straightforward to say that it is the A-intension that is the cognitive value of an expression, but on his view, it is only the A-intension to which the thinker has access.

thought is not the proposition described, but the proposition that this proposition, whatever it is, is true.

In contrasting the different interpretations of the two-dimensional framework, I have used the labels "metasemantic" (for the interpretation I want to defend) and "semantic" (for the generalized Kaplan interpretation). The terminology marks a distinction between questions about what the semantic values of expressions are and questions about what the facts are that determine those semantic values. Kaplan introduced it to contrast two different ways of understanding a causal theory of reference.[7] The direct reference account says that the semantic value of a name is simply its referent. (That is the whole *semantic* story.) The causal mechanisms explain what the facts are that make it the case that names have the semantic values that they have. (They are part of the *metasemantic* story.) A contrasting theory, "causal descriptivism," holds that the causal mechanisms belong in the semantic story: one should take the semantic value of the name to be a description something like this: "*the individual who lies at the other end of the historical chain that brought this token to me.*"[8] I labeled my interpretation "metasemantic" because the second dimension represents the facts in virtue of which the utterance in question has the semantic content that it has. The generalized Kaplan interpretation was called "semantic" because the two-dimensional intension—the analogue of Kaplanian character—and the A-intension that it determines, are semantic values of the expressions. The generalized Kaplan interpretation is a kind of generalization of causal descriptivism.

It is not important what gets called "meaning," or labeled "semantic." The significance of the contrast between the two kinds of interpretation is in the order of explanation of the fact that an utterance is associated with the particular two-dimensional intension that it is associated with, and in the kind of account of intentionality that the contrasting stories require. In the metasemantic story, the problem of intentionality is addressed at the level of C-intensions, which are the contents of thought, and the cognitive values of expressions, in the normal case. Propositional concepts are defined, for an utterance token, only relative to possible worlds in which the utterance event takes place,[9] and the diagonal propositions determined by propositional concepts are local and context-dependent. One can define a propositional concept for any context, but in the normal case, where speakers know what they are saying (according to the standard semantic rules) and hearers are presupposed to understand what is said, the propositional concept will be constant, relative to the context, and so the diagonal proposition, or A-proposition, will be

[7] See Kaplan (1989b: 574). I use this terminology to distinguish the two kinds of interpretation of the two-dimensional apparatus in Stalnaker (2001) and Stalnaker (forthcoming).

[8] Kaplan (1989: 574).

[9] As discussed in Stalnaker (1987), one extends propositional concepts to possible worlds not containing an utterance token in applications of the diagonalization strategy to belief attribution by considering what a token that-clause *would* have said if uttered in a certain possible world. But as I emphasized, the counterfactual is vague, and this is an ad hoc, case-by-case procedure that requires charitable interpretation.

the same (relative to the context) as the horizontal, or C-proposition.[10] In contrast, the generalized Kaplan interpretation addresses the problem of intentionality on the level of two-dimensional intensions, or A-intensions. Since these intensions are not defined in terms of what an utterance expresses or would express in the relevant possible world, they can be assumed to be defined for a broader range of possible worlds. What matters is not what the content of an utterance would have been if uttered in some alternative possible world, but what value the actual two-dimensional meaning takes where the argument of the function is the alternative possible world. A-intensions are, in the general case, the cognitive values of expressions, and for all cases where the C-intension depends on the external environment (including all cases involving proper names, natural kind terms, color words, or any terms to which "twin-earth" thought experiments might be constructed), the A-intension will differ from the C-intension.

The two interpretations make different assumptions about what we have cognitive access to because they have different accounts of what cognitive access is. A thoroughly externalist account of intentionality, since it gives an externalist account of thought as well as speech, gives an externalist account of cognitive access. Knowing who Socrates is, and so having cognitive access to singular propositions about Socrates, is a matter of being appropriately causally related to Socrates. Knowing who someone is, and so knowing what singular proposition is expressed by some singular statement, is of course highly context-dependent, and cognitive access will be context-dependent in the same ways. Cognitive access, on an externalist theory, is not simply a matter of the strength of an acquaintance relation. It is a matter of whether a person's state of mind is aptly described in terms of an individual (in terms of a distinction between possible worlds in which that individual has a certain property and worlds in which the individual does not).

The account of intentionality implicit in the generalized Kaplan interpretation is internalist. Two-dimensional and A-intensions are assumed to be determined, in general, by the internal properties of the speaker or thinker, and to be accessible a priori.[11] A prioricity is identified with the necessity of the A-intension—an identification that does not have any plausibility on the metasemantic interpretation. Even paradigm cases of truths knowable a priori (for example simple mathematical truths) will have contingent diagonals in some contexts, on the metasemantic account. Consider a context in which a person is uncertain about whether the intended meaning of a certain token of "$7 + 5 = 12$" is the usual one, or one that uses a base 8 notation, with the same numerals for one through seven. In some possible worlds compatible with the beliefs of this person, the token expresses the falsehood that seven plus five is

[10] To say that two propositions are the same, relative to a context, is to say that the two functions from possible worlds to truth values take the same values for all possible worlds compatible with the context. So, for example, the proposition that the current President of the United States is a Republican is the same as the proposition that G. W. Bush is a Republican, relative to a context in which it is presupposed that G. W. Bush is the President.

[11] The thesis that a sentence is a priori if and only if it has a necessary A-intension is described by Chalmers as "the core thesis" of his interpretation of the two-dimensional framework. It is made true by definition in that interpretation. See Chalmers (2002).

ten, and so the diagonal will be contingent. More generally, any utterance, no matter how trivial the proposition that it in fact is used to express, might have been used to say something false, and a person might have misunderstood it to say something false. So the metasemantic interpretation yields no account or representation of a priori truth or knowledge, and does not depend on any notion of the a priori.[12] This may be regarded as a strength or a weakness of the metalinguistic interpretation, depending on one's attitude toward the notion of a priori knowledge and truth, but it is a clear difference between the two interpretations.

3. Contextual and Epistemic Interpretations

David Chalmers is one of those I have in mind as a proponent of the generalized Kaplan interpretation, but he has a different way of contrasting his own interpretation with alternatives. Chalmers distinguishes *contextual* from *epistemic* understandings of the two-dimensional framework. The one "uses the first dimension to capture *context-dependence*," while the other uses it "to capture *epistemic dependence*."[13] This classification cuts across the semantic/metasemantic distinction that I am making, since both Kaplan's semantics for demonstratives and the metasemantic interpretation count as contextual interpretations. In Chalmers's classification, the kind of contextual interpretation that comes closest to the metasemantic account is one that identifies a two-dimensional intension with what he calls a "token reflexive contextual intension." To evaluate such an intension for a given actual utterance token at a possible world, we consider what proposition is expressed by that particular utterance token in the possible world in question. Chalmers argues that the token reflexive account is problematic since it seems to depend on questionable metaphysical assumptions about the essential properties of linguistic and mental tokens—about the way they are identified across possible worlds. To borrow and adapt an old example of Donald Davidson's, suppose Daniels says "Empedocles leaped," but O'Leary took him to be speaking German, saying "Empedocles liebt." Is the token utterance of the German sentence that Daniels utters in the possible world that O'Leary thinks we are in really the very same token as the token of the English sentence that he utters in the actual world? Do we have to assume that it is in order to use the two-dimensional framework to represent the misunderstanding? This is a good question, but I think one can bypass metaphysical questions about the essential properties of tokens. It will suffice for the metasemantic propositional concepts that the tokens in the alternative possible

[12] As David Chalmers keeps reminding me, in the face of my increasingly strident expressions of skepticism about a priori truth and knowledge, I did say, in "Assertion," that a certain two-dimensional modal operator, which says that the diagonal proposition is necessary, could be understood as the a priori truth operator (p. 85). I now think that this was an ill-considered remark. The notion of a priori truth that this identification yields is at best a very local and context-dependent one.

[13] Chalmers (2002).

worlds be epistemic counterparts of the actual token.[14] In a context to which the two-dimensional apparatus can be straightforwardly applied (either in a case where reinterpretation by diagonalization is required, or in a case of ignorance or misunderstanding of what is said), the relevant people will believe, or presuppose, that a particular utterance event takes place, so there will be a uniquely salient utterance token in each of the relevant possible worlds. So long as it is clear which utterance token it is, it does not matter whether it is literally the same one. (Though Chalmers's point does underscore the extent to which the application of the apparatus, on the metasemantic interpretation, is context-dependent. A propositional concept, and the diagonal proposition it determines, will be well-defined only for a limited range of possible worlds.)

What is the epistemic interpretation that Chalmers contrasts with all versions of the contextual theory? Here is the way I understand it: We start with an "epistemic space," a set of possibilities, or scenarios, "ways things might turn out to be, for all we know *a priori*." These scenarios can be described in a canonical language which is "semantically neutral," which means roughly that the terms in it are not "twin-earthable": the two kinds of intensions (A-intensions and C-intensions, to stay with Jackson's notation) will coincide for the terms of the canonical language. It is assumed that this special language is rich enough to give a complete description of the scenarios, or points of the epistemic space. A description is complete if knowing it would suffice to put one in a position to know any truth by reasoning alone. The two kinds of intensions are then defined as functions with the subsets of this space of possibilities as its range. Thought and speech in general are then interpreted by assigning these intensions to thought and utterance types. The project is, in effect, a project of reduction to the canonical language, for which all content is narrow, and knowable a priori.

The first thing to note about Chalmers's account of epistemic space is that since it defines the space of possibilities in epistemic terms, and takes epistemic notions such as a priori knowledge as unexplained primitives, it does not directly address what I regard as the central question of interpretation: what are the facts in virtue of which expressions are associated with the intensions (one or two-dimensional, A or C) that they are associated with? But it does put constraints on the way that question can be answered, and I am skeptical that they can be met. Since the contents of the sentences of the rich but neutral canonical language are narrow contents—determined by the intrinsic properties of the speakers of that language—it, like any version of the generalized Kaplan interpretation, needs an internalist solution to the problem of intentionality. I doubt that any such solution can be made to work, but I think one can say something about the general shape that a successful account of this kind would have to have. All attempts to address the problem of intentionality consist mainly of the waving of hands, but the different kinds of hand-waving suggest very different pictures of our intentional relations to the world. The clearest and best developed internalist account that I know of is a theory David Lewis calls "global descriptivism,"

[14] This problem is briefly noted in Stalnaker (1981: 138, n. 14).

and I think this is the kind of account that any proponents of the generalized Kaplan interpretation should find congenial.[15]

4. Global Descriptivism

Global descriptivism begins by analyzing names and some general expressions in terms of definite descriptions. The descriptions may involve causal notions, and a reference to the speaker (as in "the individual who lies at the other end of the historical chain that brought this token to me"), and the descriptions may be rigidified ("the *actual* man who corrupted Hadleyburg"). One may define several names together ("Cicero and Cataline are the men such that the first denounced the second and . . . "). But as Lewis emphasizes, an analysis of a name or a predicate in terms of a definite description simply passes the semantic buck from one part of the vocabulary of the language to another. The idea of the global theory is to interpret all the non-logical terms of a language at once. The method follows and generalizes Frank Ramsey's proposal for interpreting theoretical terms in which predicates are replaced with variables, bound by quantifiers. One does not refer directly either to individuals or to empirical properties and relations, but instead quantifies over them. The content of one's theory is that there exist properties and relations that are related to each other in the way that the laws and generalizations of one's theory say that they are. The theory is true in the set of those possible worlds that provide an appropriate model for the theory.

I said that according to global descriptivism, a speaker's theory is true in possible worlds that provide an *appropriate* model for the theory since as Lewis emphasizes, an unconstrained global descriptivism would be untenable. Properties and relations, unconstrained, are plentiful enough to provide models for any theory in any possible world, so if one required only that there be some model for the theory, then all theories would be true in all possible worlds (or at least in all possible worlds of the right size). That is Putnam's paradox, and Lewis takes it to refute an unqualified global descriptivism. The main burden of this account of intentionality, as Lewis argues, is to explain the constraints on the properties and relations that the quantifiers range over, and so that define the restricted class of models that are appropriate. Lewis's idea is that the properties and relations must be more or less natural, and the hope is that this kind of constraint will suffice to give a version of global descriptivism that gives empirical claims the kind of substantive content that they seem to have.

It will be agreed by all that this kind of theory will not be even remotely plausible unless the theory being interpreted is a rich and detailed one. Suppose one's theory said only that all swans are black. Then the global descriptivist analysis would say that the content of the theory is that there exist two properties such that everything that satisfies the first also satisfies the second. Even if the class of properties one is quantifying over is restricted to simple, natural properties, this claim will obviously not be a plausible paraphrase of the generalization about swans. To have a chance of

[15] See Lewis (1984). I discuss Lewis's global descriptivism in more detail in Stalnaker (2004c).

plausibility, the global descriptivist account must be applied to a theory with many predicates, and a large number of generalizations about the interrelations between the properties and relations that instantiate them. The individual claims that a theory makes cannot be understood independently of the whole theory.

A proponent of the generalized Kaplan interpretation of the two-dimensional framework, and of the internalist conception of content that it requires, need not subscribe to all of the details of Lewis's account of intentionality, but I think that any internalist account of intentionality will share with his account certain features that seem to me problematic. I will conclude by sketching two objections to global descriptivism, which I will call the holism problem and the indirectness problem. More neutrally, perhaps I should call them two distinctive features of that kind of theory, since others may find them less problematic than I do. In both cases, the presence of these features helps to explain the role of the second dimension in trying to reconcile this kind of account of intentionality with the phenomena of speech and thought.

First, the holism problem: Meanings and contents, on this kind of account, will be extremely unstable and idiosyncratic. Since interpretation goes by way of a total theory, any change in the total theory, however minor, will bring about a change in the contents of everything expressed in the theory, and any difference between your total theory and mine will mean that the contents of all my claims will differ from any of yours. This is a familiar objection to internalist accounts of meaning.[16] Jackson and Chalmers acknowledge that the kind of descriptivism they are defending is holistic in this way, but apparently do not take it to be a problem. It may be, they say, that "Leverrier uses 'Neptune' as a name for whatever planet perturbs the orbit of Uranus," while his wife uses the same name as a name for the "astronomical object for which her husband is searching . . . . 'Neptune (if it exists) perturbs the orbit of Uranus' is a priori for Leverrier but not for his wife." The same kind of variability might affect natural kind terms: they suggest that it might be a priori for a city dweller, but not a beach dweller, that water comes out of faucets, and a priori for the latter but not the former that water is the liquid in the ocean.[17]

The second dimension of meaning is supposed to soften the effect of the fact that we never mean the same thing as others with whom we communicate, or mean the same thing ourselves from day to day. If the descriptions that give the meanings of our names and predicates are rigidified descriptions, then "that will avoid confusion between people who have attached the same term to the same referent by means of different descriptions."[18] It is rigidification that gives rise to the general A-intension/C-intension distinction, and the derivative C-intensions will tend to be much more stable across time and person than the A-intensions from which they are derived. The C-intensions will play an essential mediating and stabilizing role. But the holism feature does yield a peculiar account of communication. It is the A-intensions that are the cognitive value of our thoughts and utterances—the

[16] Jerry Fodor, for example, takes this to be a devastating objection to conceptual role theories of meaning. See Fodor (1987).

[17] Chalmers and Jackson (2002). [18] Lewis (1984: 59).

propositions to which we have access—but they are not what must be the same for successful communication. The thought I express—what I believe when I am sincere and say what I believe—is rarely if ever the same as what you come to believe when you accept what I say.

Second, the indirectness problem: The kind of content to which we have access, according to the global descriptivist theory, is extremely abstract. It is not just that we do not refer directly to particular individuals, and entertain singular propositions. We also do not describe things in terms of ordinary empirical properties and relations, but only in terms of whatever properties and relations are the ones that best fit the abstract structure given by our uninterpreted theory. Again, it is the second dimension that is supposed to mitigate the indirectness and give us a kind of access to the individuals that inhabit our world and to the empirical properties and relations that they instantiate. Our utterances, and perhaps our thoughts, have singular propositions and propositions involving empirical properties and relations as their C-intensions. But on this interpretation, it is only the two-dimensional intensions and A-intensions to which we have cognitive access, and according to the global descriptivist account, we *never* have cognitive access to any propositions except the very abstract ones that existentially generalize over empirical properties and relations. The rigidification operations that give us the second dimension are, in effect, devices for describing propositions that we cannot grasp. We can, for example, describe the proposition that there is water (water itself, not whatever it is that plays the water role) on the floor, but that is not the proposition we believe when we believe that there is water on the floor, or the one we entertain when we consider the possibility that there is. But cannot we, who know that water is H_2O, grasp such propositions? No, because our access to Hydrogen and Oxygen is equally indirect. For the global descriptivist, it is indirect description all the way down.

Julius provides a paradigm here. In one of the early discussions of the problem we have been concerned with, Gareth Evans stipulated that the name "Julius," as he proposed to use it, should be a rigid designator for the person, whoever he or she is, who invented the zip. It seems intuitively clear—it is part of the point of the example—that some competent users of the name "Julius" (presumably including Evans himself) do not know who Julius is: they do not know to whom the name "Julius" refers. It is of course not at all clear what it takes, in general, to know who someone is, or to know what a name refers to, but this seems to be a clear case. And if someone does not know who it is that "Julius" refers to, then he does not know what singular proposition is expressed by sentences using that name, such as "Julius invented the zip," and "Julius was born in Minsk." In such a case, the person can believe that the singular proposition, whatever it is, is true, but that will not be the same as believing the singular proposition itself. So much is pretheoretical intuition, common ground whatever one's interpretation of the two-dimensional apparatus used to describe the case. It seems clear that the propositions believed by a person who is prepared to affirm the truth of these statements are descriptive propositions: the necessary truth that the inventor of the zip invented the zip, and the contingent but general proposition that the inventor of the zip was born in Minsk. These propositions are the diagonal or A-intensions of the statements.

According to global descriptivism, our access to all properties, relations, and individuals is like our access to Julius, and our relation to the C-intensions of all sentences we understand is like our relation to the contingent singular propositions that Julius invented the zip, and that Julius was born in Minsk. While we are in a sense talking about Julius when we say things using this name, we do not know what we are talking about. On a global descriptivist theory, this is the general case: we never know what we are talking about.

The metasemantic interpretation and the externalist, causal/information-theoretic account of intentionality that motivates it, agree with the generalized Kaplan interpretation with its internalist account of intentionality about the case of Julius. That is, it is common ground that it is the A-intensions that are expressed and communicated with sentences using the name Julius in the kind of context that Evans intended. In any context in which it is presupposed that "Julius" names the inventor of the zip, but in which there is no individual who is presupposed to be the inventor of the zip, diagonalization will be required. The disagreement between the different interpretations that I am contrasting is about whether one should think of this as a model for the general case.

It might be nice if we had a neutral language with an internally grounded semantics, a language that required no factual assumptions for its interpretation and that could provide a complete description of the world, and all possible worlds. It might be nice if there were a pure epistemic space to which we had a priori access and in terms of which we could locate our disagreements about what the actual world is like. But I do not think these things are possible. The only way we can describe the world is to use the materials that the actual world offers us—the things, properties and relations that we find there. Where we disagree about the nature of what is to be found in the actual world, we may as a result disagree about what is possible—about the character of the space of possibilities in terms of which our language and thought are interpreted. Semantic and factual issues become intertwined, and that is a problem. Where the disagreements are about the fundamental natures of things, or about identities, the problem is particularly acute. But our resources for describing the world are rich and diverse, and even if there is no absolutely neutral language, we can usually find ways of describing the possibilities that are neutral on the issues in contention in a particular context. The two-dimensional apparatus, on the metasemantic interpretation, is apt for describing the problems that arise from the mix of semantic and factual information, and some of our resources for solving them.

References

Chalmers, David (1996). *The Conscious Mind*. Oxford: Oxford University Press.

——(2002). "The Foundations of Two-dimensional Semantics." (on line paper, www.u.arizona.edu/~chalmers/papers/foundations.html)

——and Frank Jackson (2001). "Conceptual Analysis and Reductive Explanation," *The Philosophical Review* 110, 315–61.

Fodor, Jerry (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: Bradford Books, the MIT Press.

Jackson, Frank (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

——(2001). "Précis of *From Metaphysics to Ethics*," *Philosophy and Phenomenological Research*, 62, 617–24.

Kaplan, David (1989a). "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford and New York: Oxford University Press.

——(1989b). "Afterthoughts," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford and New York: Oxford University Press.

Lewis, David (1984). "Putnam's Paradox," *Australasian Journal of Philosophy*, 62, 221–36. (Reprinted in Lewis, *Papers in Metaphysics and Epistemology* (Cambridge Studies in Philosophy, Cambridge: Cambridge University Press, 1999). Page numbers refer to the latter.)

Stalnaker, R. (1978). "Assertion," *Syntax and Semantics*, 9, 315–32. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1981a). "Indexical Belief." *Synthese*, 49, 129–51. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1981b). "Logical Semiotic." E. Agazzio (ed.), *Modern Logic—A Survey*, Dordrecht: D. Reidel, 439–56.

——(1987). "Semantics for Belief." *Philosophical Topics*, 15, 177–90. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1999). *Context and Content*. Oxford: Oxford University Press.

——(2001). "On Considering a Possible World as Actual," *Proceedings of the Aristotelian Society* (supplementary volume). Reprinted in Stalnaker (2004a), 188–200.

——(2004a). *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*. Oxford: Oxford University Press.

——(2004b). "Conceptual Truth and Metaphysical Necessity." In Stalnaker (2004a), 201–15.

——(2004c). "Lewis on Intentionality," *Australasian Journal of Philosophy* 82, 199–212. Reprinted in F. Jackson and G. Priest, *Lewisian Themes: The Philosophy of David Lewis*. Oxford, Oxford University Press, 2004, 231–44.

14

Two-Dimensionalism and Kripkean A Posteriori Necessity

Kai-Yee Wong

Three decades on, despite wide acceptance of Kripke's examples of necessary a posteriori truths, the question of how to explain such truths is still very much alive. A notable, extremely promising approach is the *two-dimensional* strategy, which derives from various writings on 'Naming and Necessity' (Kripke 1972) by Martin Davies and Lloyd Humberstone (1980), David Kaplan (1989), David Lewis (1981), Robert Stalnaker (1978), and, most recently and explicitly, David Chalmers (1996, 2004), Frank Jackson (1998) and Kai-Yee Wong (1990, 1996a). Yet, as I will argue, most proponents of the two-dimensional approach seem unaware of, or have paid inadequate attention to, a serious objection, one that threatens to undermine not only their particular arguments, but the very idea of a two-dimensional explanation of a posteriori necessity. In this essay, I will explain this objection, the *dual-proposition problem*, by explicating the *associated-proposition strategy* that underpins the explanations proposed by Jackson and Chalmers. I will also indicate how I think two-dimensionalists can best respond to this threat.

The essence of the associated-proposition strategy is to distinguish the necessary proposition *expressed by* a sentence—say, 'Water is H_2O'—from the a posteriori proposition *associated with* the sentence. This strategy lies behind a number of criticisms and explications of Kripke's contention that there is such a thing as a posteriori necessity. The distinctive feature of the two-dimensional approach is that it provides an abstract, double-index framework that represents the deep, underlying relationship between the two sorts of propositions. Section 1 of this paper outlines a version of this framework that I previously proposed based on Stalnaker's work. Before presenting the dual-proposition problem, I summarize in Section 2 the two-dimensional explanations of the necessary a posteriori offered by Jackson and Chalmers, explaining the core ideas they share, in particular the associated proposition strategy. Once this strategy is made explicit, one can see how their views are subject to the dual-proposition objection. Or so I argue in Section 3, where I set forth the dual-proposition objection and explain the threat

I am grateful for advice from Manuel García-Carpintero, Josep Macia, two anonymous referees for Oxford University Press, and Chris Fraser.

it poses to the two-dimensional approach. I then turn to Pavel Tichy's version of the dual-proposition objection in order to reveal a hidden, crucial component of the objection: the assumption that 'a priori' and 'a posteriori' apply, in the first instance, to *propositions* and do so *simpliciter*. I call this view the *absolute view* of propositional a priority, which I contrast with the *relative view*. In the final, fourth section I explain how the dual-proposition objection relies on the absolute view, and then how two-dimensionalists can avoid this objection by adopting the relative view. I conclude the essay by offering some general remarks that put the need to relativize propositional a priority into a wider perspective.

1. A Two-Dimensional Framework

From the early 1970s through the early 1980s, constraints on semantic content posed by facts about context of use and the influence of context on the determination of truth values constituted central topics in such areas as formal pragmatics, context-sensitive semantics, logical pragmatics, double-index semantics, and two-dimensional modal logic.[1] The approach to the necessary a posteriori discussed in this essay is an application of an abstract two-dimensional semantic framework that is a synthesis of results from investigations in these areas.

Among the two-dimensional frameworks that have been proposed, Robert Stalnaker's logical-pragmatic apparatus of two-dimensional matrices distinguishes itself by its transparent representation of two-dimensionality in terms of *diagonalization* (Stalnaker 1972, 1978, 1999). What follows is an outline of a Stalnakerian two-dimensional framework.[2]

We learned from Kripke that 'water' is a rigid designator for the substance that possesses such superficial properties as being colorless, tasteless, the liquid that fills our lakes, and so forth—in short, the property of being the watery stuff. Suppose 'water' was introduced by way of a reference-fixing description in terms of watery properties. Because the watery stuff in the actual world, W_1, is H_2O, 'water', being rigid, refers to H_2O with respect to Twin Earth, W_2, in which the watery stuff is XYZ. Thus, if Kripke is right, the statement

(1) Water is H_2O

expresses a necessary proposition, which can be represented by the top (or equivalently the bottom) row of the matrix below (assuming there are just a small number of worlds). Yet under the supposition that the actual world, that is, the actual context in which (1) is asserted, is W_2, the statement expresses a different, necessarily false proposition, because what satisfies the relevant reference-fixing description in W_2 is not H_2O but XYZ. If we suppose that W_3 is an H_2O-world, then the way

[1] See Åqvist (1973); Bar-Hillel (1954); Hansson (1974); Kaplan (1989); Kamp (1971); Lewis (1972) and (1981); Montague (1970); and Stalnaker (1978), (1980), (1981).

[2] What is presented here is a simplified version of the framework I develop in Wong (1990) and (1996a).

the propositional content of 'Water is H_2O' depends on the way the world is can be represented thus:

	W_1	W_2	W_3
W_1	T	T	T
W_2	F	F	F
W_3	T	T	T

A

Matrix **A** is what Stalnaker calls the *propositional concept* determined by (1). It is a function taking a world to a proposition (represented as a set of possible worlds), or, equivalently, a function taking a pair of worlds to a truth-value. (For a singular term, one may take the relevant function as assigning to each pair of worlds an object of the relevant sort.) **A** gives a representation of the two different ways in which facts determine the truth value of what is said by an utterance. These two ways correspond to the different roles possible worlds play in the matrix. The vertical axis represents possible worlds in their role as *context*—as what determines what is said. I have called the possible worlds playing such a role *context-worlds*.[3] The horizontal axis represents possible worlds in their role as *evaluation circumstance*—as what determines whether what is said in a certain possible world considered as the actual context is true. The distinction between these two roles is often conveniently expressed as the distinction between a world considered as the *actual world* and a world considered as a *counterfactual situation*.

With these preliminaries, we can now turn to the important concept of *diagonalization*. From the two-dimensional point of view, the points on the *diagonal* (from top left to bottom right) of a two-dimensional matrix have a unique and crucial theoretical role to play, in that they represent where the context-world is identical with the world of evaluation. I have called these points *good-points*. We can now distinguish two kinds of extensions for any expression *E*. The *one-dimensional extension*, or extension in the traditional sense, of *E* with respect to a world *w* is the semantic value of *E* in *w* with *the actual world* considered as the context-world. The *two-dimensional extension* of *E* with respect to *w* is the semantic value of *E* in *w* with *w* considered as the context-world—in other words, the value of *E* in the good-point <w, w> of the relevant matrix.[4] For example, since the watery stuff in the actual world is H_2O,

[3] A more general notion is what Quine (1969) calls a 'centered possible world', which is an ordered pair consisting of a possible world and a center (consisting of a time and an individual in the world). The center is necessary for cases involving indexical terms such as 'I'. See also Chalmers (1996): 60–1.

[4] $<i, j>$ represents a point in a matrix, where *i* is the relevant world on the horizontal axis and *j* the relevant world on the vertical axis.

the one-dimensional extension of the rigid designator 'water' with respect to any *w* is H_2O; but the two-dimensional extension of the term in any world *w* is the watery stuff in that world. Correspondingly, we can distinguish two kinds of intensions. The *one-dimensional intension* of *E* is the function assigning to each possible world the one-dimensional extension in that world. The *two-dimensional intension* of *E* is the function assigning to each possible world the two-dimensional extension of *E* in that world.[5] Accordingly, the one-dimensional intension of 'water' is the constant function taking each world to H_2O, and the two-dimensional intension of 'water' assigns H_2O to the actual world, XYZ to the XYZ-world, PQR to the PQR-world, and so on. For a sentence, the one-dimensional intension is a proposition in the traditional sense, and the two-dimensional intension is the diagonal proposition of the relevant propositional concept.

To determine whether a sentence as used in a context-world *w* is necessary, we need consider only the proposition expressed by that sentence in that world. But from the two-dimensional point of view, to determine whether the sentence is a priori or a posteriori, we must look at the relevant propositional concept, or more specifically, the relevant diagonal proposition. I have called a propositional concept *quasi-necessary* if its diagonal proposition is necessary and *quasi-contingent* if its diagonal proposition is contingent.[6] The essence of the present account is the suggestion that Kripkean a posteriori necessity arises just when the proposition expressed by a sentence (in the actual world) is necessary, but the propositional concept determined by the sentence is quasi-contingent.

As presented above, two-dimensionalism gives prominence to the distinction between a world considered as the actual world and a world considered as an evaluation circumstance, with the notion of the actual world construed *contextually* (that is, as a world in the role of context of utterance).[7] The notion of an evaluation circumstance, on the other hand, has traditionally been associated with contingency,

[5] My use of 'two-dimensional intension' here differs from the terminology of Stalnaker (2001) and Chalmers (2004). In their terminology, a two-dimensional intension is a function taking a world to a one-dimensional intension, that is, a function taking two arguments, a context-world and a circumstance of evaluation, to a semantic value of the relevant kind. I give 'two-dimensional intension' a narrower reading by requiring the two arguments to be identical. In general, a two-dimensional intension in this narrow sense is the result of diagonalizing a two-dimensional intension as understood by Stalnaker and Chalmers. What Jackson calls an *A-intension* is two-dimensional in this narrow sense (see Section 2 below).

[6] For a justification of explicating a priority in terms of quasi-necessity, see Wong (1996a): 73–80.

[7] But Chalmers thinks that the contextual construal is fundamentally mistaken. Chalmers, who once wrote that 'when we consider a world *w* as actual, we think of it as a potential context of utterance, and wonder how things would be if the context of the expression turned out to be *w*' (Chalmers 1996: 60), has recently come to think that 'a world considered as actual' should be given a substantially, or even fundamentally, different reading—that is, an *epistemic reading*—in order to distinguish it from 'a world considered as a context'. Otherwise, Chalmers (2004, see also forthcoming) argues, the two-dimensional account will not yield correct results about the epistemic status of such sentences as 'Words exist' and 'Language exists', which are true whenever uttered. (It should be noted that a similar kind of sentence has drawn the attention of a number of theorists in recent years. Some, notably Kaplan (1979, 1989), argue that 'I am here now', being a logical truth, is

in that a proposition may be true in some evaluation circumstances but not others, or with necessity, in that a proposition may be true in all circumstances. The unique insight of two-dimensionalism, as I see it, lies not so much in drawing this distinction as in the subtle way it ties the notion of the actual world construed contextually to the notion of an evaluation circumstance (and in turn to that of contingency) by way of diagonalization.[8] This tie is the source of—and as such provides the basis for a two-dimensional explication of—the apparent contingency of Kripkean a posteriori necessary truths.

2. Jackson's and Chalmers's Two-Dimensional Explanations

Recently Jackson (1998) and Chalmers (1996) have independently proposed two-dimensional accounts of Kripkean a posteriori necessary truths. Their accounts, as well as the one I proposed earlier (Wong 1990, 1996a), can be regarded as variants of a general approach naturally suggested by two-dimensional logic.

Jackson (1998: 48) distinguishes what he calls the A-extension/intension and the C-extension/intension of a term ('A' for actual and 'C' for counterfactual). The *A-extension* of a term, for each world *w*, is 'what the term applies to in *w*, given or under the supposition that *w* is the actual world, our world'. The *C-extension* of a term T, for each world *w*, is what T applies to in *w* 'given whatever world is in fact the actual world'—in other words, the extension of T in a counterfactual world. Jackson calls

a priori. Gerald Vision's (1985) argument that the standard telephone answering-machine message 'I am not here now' provides a counterexample to Kaplan's view generated an interesting exchange in the pages of *Analysis*. See Colterjohn and MacIntosh (1987), Simpson (1987), and Vision (1987). Salmon (1991) argues that Vision's example is 'best thought of as a genuine case of *assertion in absentia*'. Against this, I (1996b) argue that the sentence may be informative in some particular contexts only because the hearer can pragmatically exploit the fact that in its normal use it is patently a logical and thus trivial falsehood.)

[8] Arguably, diagonalization is not only a central notion in two-dimensionalism but also a concept close to the foundation of formal semantics. To appreciate this point, one may observe that the duality (i.e., the dual role of a possible world) that diagonalization is meant to capture is also reflected in the two different ways in which 'the actual world' can be obtained in formal semantics (see Kaplan 1989: 594–6). The first way is, as Kaplan says, 'by starting with a full-blown indexical language, deriving the notion of context from its role in the semantics of indexicals, and then recognizing that truth, absolute truth in a model, is assessed at the world-of-the-context, i.e., the actual world' (1989: 595). The intuitive idea behind this is that the world in which a context occurs is the same world that is *actual* from the point of view of the context in question. Alternatively, for an indexical-free modal language, (absolute) truth (in a model) is truth in the *designated* world (of the model). Intuitively, the designated world is the actual world. This can be shown by the usual interpretation of the 'actually true' operator, if it is to be added to the language: relative to a model, *s* is *actually true* with respect to a world *w* if and only if *s* is true with respect to the designated world. So the second way of obtaining the notion of the actual world is to start with an indexical-free language and recognize that absolute truth in a model is *evaluated* in the designated world. One might think that this notion of the actual world is different from that obtained in the first way. But this is not true. The designated world is what remains if one takes away all contextual parameters save the world-of-context. 'The actual world' obtained in this second way may thus be regarded as the notion of context in the limiting sense, in other words, a residue of the notion.

the function assigning to each world the A-extension of a term in that world the *A-intension* of the term. The function assigning to each world the C-extension of a term in that world is the *C-intension* of the term. Accordingly, the *A-proposition* of a sentence is 'the set of worlds satisfying the following condition: given that *w* is the actual world, then the sentence is true at *w*'. The *C-proposition*, the 'one we have been calling the proposition expressed', is 'the set of worlds at which the sentence is true given which world is in fact the actual world' (Jackson 1998: 76).

Jackson (1998: 73–4) then distinguishes two senses of 'knowing the conditions under which a sentence is true'. The propositions expressed by our 'water' sentences depend on how things are in our world. Anyone who does not know that the watery stuff of our acquaintance is H_2O does not know the truth conditions under which 'Water covers most of the earth' is true, in the sense that they 'could know all there is to know about some counterfactual world without knowing whether the sentence is true in that world . . . through their ignorance about the actual world'. Nevertheless, they must know the truth conditions of 'Water covers most of the earth' in the sense that they know 'how the proposition expressed depends on context of utterance—in this case, how it depends on which stuff in the world of utterance is the watery stuff of our acquaintance in it'.

It is not difficult to see that the truth conditions involved in the first case relate to the C-proposition and in the second case the A-proposition. While the former is what is normally meant by the 'unadorned use' of the 'proposition expressed by a sentence', it is 'the A-proposition we know in virtue of understanding a sentence' (Jackson 1998: 76).

Given these distinctions, Jackson claims, the explanation of the necessary a posteriori is now straightforward. 'Water is H_2O' is necessary because its C-proposition is necessary. But understanding it requires only knowing the A-proposition. Therefore one can understand the sentence without knowing enough to see that the sentence is necessary or even that it is true. (See Jackson 1998: 77.)

Similarly, Chalmers thinks 'a two-dimensional picture of meaning and necessity' provides 'a natural way of capturing Kripke's insights'. He starts with a general observation characteristic of a two-dimensionalist:

> There are two quite distinct patterns of dependence of the reference of a concept on the state of the world. First, there is the dependence by which reference is fixed in the *actual* world, depending on how the world turns out: if it turns out one way, a concept will pick out one thing, but if it turns out another way, the concept will pick out something else. Second, there is the dependence by which reference in *counterfactual* worlds is determined, given that reference in the actual is already fixed. (Chalmers 1996: 57)

Corresponding to these two types of dependence, Chalmers (1996: 57–8) distinguishes the *primary* intension and the *secondary* intension of a concept. The primary intension maps worlds to extensions 'reflecting the way that actual-world reference is fixed'. So 'the primary intension of "water" maps the XYZ-world to XYZ, and the H_2O-world to H_2O, . . . or more briefly, it picks out the watery stuff in a world'. Unlike the primary intension, which '*specifies* how reference depends on the way the external world turns out' and so 'does not itself depend on the way the external world

turns out', the secondary intension of a concept is not determined a priori. In the case of rigid designators such as 'water', its secondary intension maps a world *w* to the result of evaluating the relevant reference-fixing description in the actual world.

According to Chalmers (1996: 63–4), the *primary* intension is 'most central' in explaining the necessary a posteriori. He calls the primary intension of a sentence the *primary proposition associated* with the sentence. The crux of his explanation is to note the variety of *necessity* construed as 'truth across possible worlds, as long as these possible worlds are construed as contexts of utterance'. A statement is necessarily true in this ('a priori', in Chalmers's words) sense if 'the associated primary proposition holds in all centered possible worlds (that is, if the statement is necessarily true in any context of utterance)'.[9] The other variety of necessity, corresponding to the 'more familiar superficial necessity', is defined in terms of the associated secondary proposition being true in all counterfactual worlds.

Since the primary intensions of 'water' and 'H_2O' differ, the primary proposition associated with 'Water is H_2O' holds only in those centered worlds in which the watery stuff has a certain molecular structure, and thus is not *necessary* (in the sense of the first variety of necessity above). So, we cannot know on a priori grounds that water is H_2O. The secondary intensions of the two terms, however, coincide. So 'Water is H_2O', though a posteriori, is necessary, because its associated secondary proposition is necessary. This provides an account of the necessary a posteriori.

It is not difficult to see that the three accounts I have described share the same core ideas. Where my own account emphasizes the central importance of good-points, Chalmers and Jackson emphasize the crucial role of primary intensions or A-intensions. My notion of 'quasi-necessity' is translatable into Chalmers's 'first variety of necessity'. The way I handle a priority in terms of quasi-necessity echoes Jackson's discussion of the second sense of knowing the conditions under which a sentence is true. The corresponding idea in Chalmers's account is his discussion of how primary intensions are independent of empirical factors. Most important, all three accounts hold that most problems arising from Kripke's discussion of the necessary a posteriori are consequences of the unavailability of double-indexing or two-dimensional elements in the traditional conception of propositions and necessity.

3. The Dual-Proposition Problem

In discussions of Kripkean a posteriori necessity, there has been a tendency to talk in terms of a true sentence (or statement) being necessary a posteriori, even when it is by appealing to properties of propositions (or intensions) that one is attributing the property of being 'necessary' or 'a posteriori' to the sentence. For instance, Jackson holds that 'it is sentences, or if you like statements or stories or accounts in the sense of assertoric sentences in some possible language, that are necessary a posteriori' (1998: 71). Nevertheless, Jackson rejects the view that 'necessity and possibility are at bottom properties of sentences' (1998: 80). I second Jackson's rejection of this view. For those of us who do not object to talk of propositions as sets of possible worlds, it would be

[9] Chalmers's use of 'centered possible worlds' is due to Quine; see note 3 above.

self-defeating not to. For it is precisely because we want to explicate necessary 'sentences' in terms of the properties of *propositions* that we engage in proposition talk. It is important to note that two-dimensionalists characteristically reject an analogous, 'sentential' view regarding 'a priori' and 'a posteriori', at least when explicating the necessary a posteriori. As we have seen, a priority, according to Jackson or Chalmers, is at bottom a matter concerning A-*propositions* or associated primary *propositions*.

The point of positing propositions is to enable us to abstract from sentences, whose truth may vary from time to time, interpretation to interpretation, or language to language. So propositions are the *primary bearers* of truth. I take it that this is uncontroversial, especially for those who regard propositions as sets of possible worlds. For them, whether a sentence is true is a matter of whether the proposition expressed determines a set of worlds that includes the actual world.

Propositions are also regarded as objects of knowledge and belief. This view is not uncontroversial, but it is uncontroversial enough among those who seek to explicate the necessary a posteriori in a two-dimensional way. 'Objects' here need not be propositions *expressed* by sentences, of course, but they are nevertheless propositions *associated* with sentences in one way or another. In Jackson's view, which he thinks is also the view of many others, such propositions are A-propositions:

> It is, as Stalnaker, Tichy, and Chalmers emphasize, the A-proposition expressed by a sentence that is often best for capturing what someone believes when they use the sentence. . . . (Jackson 1998: 76)

Yet, combined together, the observations in the last three paragraphs yield what I will argue is a serious problem. Once we admit that necessity and a posteriority as properties of sentences are derived from properties of propositions, the possibility arises that the sentential properties in question, for a given necessary a posteriori sentence, are not derived from *one and the same proposition*. In a two-dimensional framework, this is not a mere possibility. If we take a one-dimensional proposition and a two-dimensional proposition, it may turn out that what we have is in fact one and the same proposition (for both propositions are sets of possible worlds). This, however, cannot be the case with the two propositions connected with a necessary a posteriori sentence. The two-dimensional strategy is, as it were, one of *divide and rule*: The claim that a certain sentence is necessary a posteriori is divided into two, a modal claim backed by the necessary one-dimensional proposition expressed and an epistemic claim backed by the contingent two-dimensional proposition *associated* with the sentence. *Two* propositions are involved, each of them bearing half of the burden of the claim that the sentence is necessary a posteriori. Jackson is explicit about this *dual-proposition*, or *associated-proposition*, strategy in his account:

> [T]here are two propositions connected with a sentence like 'Water is H_2O', and the sentence counts as necessary if the C-proposition is necessary, but as understanding the sentence only requires knowing the A-proposition, little wonder that understanding alone is not enough to see that the sentence is necessary. (Jackson 1998: 77)

> [W]e could say, following Tichy, Chalmers, Lewis, and Stalnaker among others, that there are two propositions connected with a sentence like 'Water covers most of the Earth'. (Jackson 1998: 76)

Similarly, Chalmers writes:

> Kripkean a posteriori necessity arises just when the secondary intensions in a statement back a necessary proposition, but the primary intensions do not. (Chalmers 1996: 64)

The question here is whether or not we have a *single* necessary a posteriori truth. One might contend that we do: it is the true sentence in question. But this answer does not square with the above observation about propositions being the primary truth-bearers and objects of belief. For at bottom what we have, on a two-dimensionalist account, are *two* propositions, a necessarily true proposition and a distinct proposition known a posteriori. To put the point another way, if someone insists that 'true' applies in the first instance to propositions, then he can object that we do not yet have a single proposition to which we can ascribe *both* necessity and a posteriority. The proposition expressed by 'Water is H_2O' is necessary, but it is not a posteriori, or else there would be no need for an associated proposition, such as the A-proposition, or the primary intensions, associated with the sentence. Conversely, the associated proposition, even if a posteriori, cannot serve as an example of a necessary a posteriori true proposition for the obvious reason that it is *contingent.*[10]

This problem is not new. It is closely related to an objection against Kripkean a posteriori necessity raised by Tichy (1983). Tichy distinguishes between the *proposition expressed by* a sentence S in language L and the *proposition associated with* S in L, where the former proposition is 'whatever (if anything) S says in L' and the latter 'the proposition to the effect that S is true in L' (Tichy 1983: 231). Accordingly, the proposition associated with

(2) Hesperus is Phosphorus

is:

'Hesperus is Phosphorus' is a true sentence in English.

Tichy then argues that

> The only way to make sense of Kripke's argument is by assuming that when he insists that [(2)] is a posteriori he does not mean that what [(2)] says can only be known a posteriori. He is not ascribing a posteriority to the proposition *expressed* by the sentence but rather to the proposition *associated* with it . . . But if this is what Kripke means, his argument is powerless

[10] Some might think that a defender of the two-dimensional strategy could consider the option of regarding the two-dimensional matrix as the 'proposition' (as suggested by an anonymous referee). But this suggestion has, I think, little to recommend it. First, a two-dimensional matrix has to be constructed out of 'propositions' in the usual sense. Second, this suggestion means giving up the 'sets-of-possible-worlds' conception of propositions underlying all the two-dimensional proposals we have considered. Third, it is not clear how one might assign a truth value (for what is said by a sentence) to a 'proposition' construed as a matrix. In other words, these considerations show that it is by no means clear how a proposition so construed can properly be called a 'proposition'. Finally, as already noted, the two-dimensional approach is guided by the idea that, while necessity is a property of *the proposition expressed* by a sentence, the epistemic status of a sentence is underdetermined by any such property and should rather be made sensitive to other, two-dimensional properties of the sentence. So, conflating the proposition with the matrix, even if it made sense, would deprive the approach of its distinctive appeal.

to cast doubt on the coextensiveness thesis [that is, that the class of necessary truths coincides with the class of a priori truths]. (Tichy 1983: 233)

The 'two-propositions interpretation' that Tichy believes is the only way to make sense of Kripke's argument is a version of a fairly common defense of Kripkean necessary a posteriority. Alvin Plantinga (1974) also supports a version of it.[11] The proposition associated with (2) that he proposes is (where Q is the proposition expressed by both 'Hesperus is Hesperus' and 'Hesperus is Phosphorus'):

'Hesperus is Phosphorus' expresses the proposition Q.

Tichy and Plantinga both hold that the proposition associated with (2) is meta-linguistic, albeit in different ways. For Chalmers, Jackson, and Wong, on the other hand, the associated proposition is not meta-linguistic, at least not explicitly so. Their two-dimensional, possible-world accounts enable them to identify *directly*, for each sentence in question, an associated proposition in terms of a set of possible worlds, without recourse to a corresponding meta-linguistic sentence. Of course, one may still ask whether there nevertheless is a meta-linguistic sentence corresponding to each such set of worlds, so that the two-dimensional strategy is only a meta-linguistic one in disguise. This may be an interesting question, but we need not be detained by it here. For what distinguishes two-dimensional accounts is the way they apply a general, abstract framework to represent the deep, underlying relationship between the two sorts of propositions connected with an a posteriori necessary sentence. If it turns out that every relevant associated proposition has a meta-linguistic sentence that expresses it, this means only that we have, in addition, a two-dimensional representation of how some sentences are related to corresponding meta-linguistic ones.

For our purposes, what is important is not how the proposition associated with a purported a posteriori necessary sentence is to be characterized. Rather, it is that Tichy has shown that the 'associated-proposition' approach will not yield any *single* proposition to which one can ascribe *both* necessity and a posteriority. I take it that the two-dimensional explanation of the necessary a posteriori as exemplified by Jackson and Chalmers, among others, is essentially an associated-proposition approach and thus subject to Tichy's kind of objection.

4. Absolute and Relative A Priority

The two-dimensional approach can be modified to meet the above objection, I suggest, by incorporating a *relative* account of the a priori. We can explain what such an account is by considering an assumption behind Tichy's interpretation of Kripke's claims about a posteriori necessity. This widely shared assumption holds that 'a priori' and 'a posteriori' apply primarily, and in the first instance, to *propositions* and do so *simpliciter*.[12] Let us call this assumption *the absolute view*, which is

[11] See Plantinga (1974): 81–7.

[12] I discuss this assumption in Wong (1996a), in which I call it 'assumption (T)'. Michael identifies this assumption in Michael (1998: 119–20). See also Salmon (1993).

both the traditional and the predominant view of propositional a priority and a posteriority.[13] The absolute view is essential to the dual-proposition objection, for given the absolute view, propositions are things to which 'a posteriori' and 'a priori' apply *directly*. That is, a proposition cannot be a priori through one mode of access but a posteriori through another. If it is a priori (or a posteriori), then it is so *simpliciter*, or non-relatively. Now if 'Hesperus is Hesperus' is a priori because it expresses an a priori proposition, then what Kripke claims is a posteriori cannot be 'Hesperus is Phosphorus', for it expresses the same, thus a priori, proposition as 'Hesperus is Hesperus'. The only plausible interpretation is that what Kripke takes to be a posteriori is a *different*, associated proposition. But in that case what backs the claim that 'Hesperus is Phosphorus' expresses a necessary proposition may no longer back the claim that this other, a posteriori proposition is necessary too. And in fact it no longer does, as Tichy argues. So we have two propositions—a necessary a priori proposition, on the one hand, and a contingent a posteriori proposition, on the other. This is the dual-proposition problem.

But nowhere does Kripke commit himself to the absolute view. Though both Jackson and Chalmers think that two-dimensionalism is 'implicit' in his writings,[14] Kripke distinguishes himself from most two-dimensionalists he has inspired by his reluctance to talk in terms of propositions. The term 'proposition' hardly occurs in the text of 'Naming and Necessity', save in the 1980 preface to the book edition, where Kripke (1972: 21) briefly replies to some of his critics and declares, 'I am unsure that the apparatus of "propositions" does not break down in this area'. And the issue that prompted him to express this concern was precisely the general issue of 'how to treat names in epistemic contexts'.[15]

By contrast, Jackson and Chalmers are clearly committed to the absolute view. On their accounts, a sentence can be a priori or a posteriori only in a derivative sense, depending on whether the relevant associated proposition (the A-proposition or associated primary proposition) is necessary or contingent. That is, propositional a priority and a posteriority, in their accounts, *are modeled on propositional necessity and contingency*, which, according to their possible-world framework of propositions, are clearly understood in a *simpliciter* or absolute sense. For it does not make sense to say that a proposition, as a function from worlds to truth-values, is necessary relative to one thing but contingent relative to another. A commitment to the absolute view is therefore part and parcel of the kind of account Jackson and Chalmers have offered. Hence, though Tichy's dual-proposition objection may be considered questionable by those who do not share his absolute view, it clearly cannot be so considered by Jackson or Chalmers.

[13] By 'propositional necessity' and 'propositional a priority' I mean, respectively, the property of necessity and the property of a priority as applied to propositions. The notions are neutral regarding whether these properties should be applied *simpliciter* or relatively.

[14] Chalmers thinks that two-dimensionalism provides 'a natural way of capturing Kripke's insights' (1996: 56). Jackson remarks that the distinction between A-intensions and C-intensions is 'implicit' in Kripke's writings (1998: 47).

[15] See also Kripke (1979).

Since the absolute view stands behind the dual-proposition argument, there is a straightforward solution to the dual-proposition problem that preserves the insights of the two-dimensional approach. The solution is to go *relative* on propositional a priority and a posteriority.

The idea of *relativizing* the epistemic status of a proposition has been entertained by Kripke himself,[16] but so far as I know, it was first suggested in publication in 1983, as a note in an essay by Keith Donnellan:

> If we distinguish a sentence from the proposition it expresses, then the terms 'truth' and 'necessity' apply to the proposition expressed by a sentence, while the terms 'a priori' and 'a posteriori' are sentence relative. Given that it is true that Cicero is Tully (and whatever we need about what the relevant sentences express), 'Cicero is Cicero' and 'Cicero is Tully' express the same proposition. And the *proposition* is necessarily true. But looking at the proposition through the lens of the *sentence* 'Cicero is Cicero', the proposition can be seen a priori to be true, but through 'Cicero is Tully' one may need an a posteriori investigation. (Donnellan 1983: 88, note 2)

Versions of the relative view have been argued for or discussed by Wong (1990, 1996a, 1996b), Michaelis Michael (1998), Heimir Geirsson (1994), and Nathan Salmon (1991, 1993). As a first approximation,[17] the relative view holds that

> A proposition *p* is a priori relative to a sentence S that expresses it if and only if S is a priori; *p* is a posteriori relative to a sentence S′ that expresses it if and only if S′ is a posteriori,

where a sentence is a priori (or a posteriori) depending on whether it determines a necessary (or contingent) two-dimensional proposition (such as an A-proposition or associated primary proposition).

Some may want to replace 'a sentence S' by something like 'a way of taking *p*' or 'a mode of access to *p*'. Indeed, a major task in elaborating the relative view is to answer the question, 'What is it that a proposition can be said to be a priori relative to?' Here I will not take a position on this question. Instead, I shall devote the rest of my discussion to explaining how the relative view resolves the dual-proposition problem. I will then conclude with some general remarks explaining why there is a need to relativize propositional a priority and a posteriority.

1. As already shown, the dual-proposition argument assumes that a genuine example of an a posteriori necessary statement must be a necessary proposition to which the concept 'a posteriori' applies in an absolute, *simpliciter* sense. The relative view rejects that assumption. On the relative two-dimensional account I have suggested, a genuine example of a necessary a posteriori *sentence* is an a posteriori sentence (in the sense defined above) that expresses a necessary proposition, and a genuine example of a necessary a posteriori *proposition* is specifiable only relative to some sentence. In other words, the reason the relative account is immune to the dual-proposition argument is *not* that we have successfully identified some necessary

[16] As reported in Salmon (1991).

[17] And following the *sentence*-relative version I argue for in Wong (1990) and (1996a).

propositions to which we can ascribe a posteriority non-relatively. Rather, it is that we simply do not accept the notion of a proposition being a posteriori or a priori in an absolute sense. It does not make sense to ask, from the relative point of view, whether the proposition expressed by 'Water is H_2O' or 'Hesperus is Phosphorus' is a priori or a posteriori *simpliciter*. The proposition can be said to be a priori or a posteriori only relative to certain particular sentences or statements. Moreover, it is important to note that the relative account is no less two-dimensional than Jackson's or Chalmers's, because it defines, as we have seen, 'an a priori (or a posteriori) sentence' in two-dimensional terms. Hence the insights of two-dimensionalism are preserved.

2. The relevance of the relative proposal is not limited to the dual-proposition problem, nor is the proposal merely an ad hoc solution to the problem. I can clarify this point by means of the following example (adopted from Michael 1998), which seems to show that any true proposition can be expressed by a sentence whose truth is known a priori. Consider the contingently true sentence 'Mary was born in Seattle' and a newly introduced sentence, '##'. The semantics of '##' is as follows: '##' expresses the same proposition as 'Mary was born in Seattle' if Mary was born in Seattle; '##' expresses the same proposition as 'It is not the case that Mary was born in Seattle' otherwise. Now let me assert that ##. I know that what I have asserted expresses a truth, a contingent truth, and I know that a priori.

This is clearly a trick, as Michael points out, but a trick that works. A natural response to it would be to point out that we know that '##' expresses a truth but do not know what truth that is. This response involves the complex issue of what it takes to grasp the meaning of a sentence. Michael plausibly argues that 'the claim that I do not know which proposition is expressed by "##" cannot be spelt out in a manner that has a principled ground' (Michael 1988: 122). I cannot take up this issue here. Instead, I want to point out that from the relative point of view, there is nothing particularly surprising about the claim that the trick works. The semantic content of '##' is contrived in such way that the sentence expresses the same truth as 'Mary was born in Seattle' but differs from the latter in being true in any context of use. That is, we have here a contrast analogous to that between 'Peter is at place l at time t' and 'I am here now' as uttered by Peter in some appropriate context. Like 'I am here now', what '##' expresses is in a sense a priori: the proposition expressed by '##' is a priori *relative to* the way we present it through the sentence '##'. But the same proposition, presented through 'Mary was born in Seattle', remains a posteriori relative to that presentation. So, many philosophers' prickly reaction to the claim that we know that ## a priori can be assuaged if one is brought to look at this example from a relative point of view.[18]

3. Yet is it not true that there is an almost universal tendency to say, in cases such as '##' or 'Cicero is Cicero', that we know a priori that ## or that Cicero is Cicero and leave it at that, without relativizing the proposition expressed to '##' or 'Cicero is Cicero'? Do we not still have this tendency even when we have come to see things from the 'relative' point of view? Here it is helpful to note that the relative view need not entail that every ascription of a priority or a posteriority must

[18] See also Michael's (1998: 121–2) discussion of the 'prickly reactions'.

be *explicitly* relativized. When taken in a certain way, such *non-relative* constructions as 'The proposition that so and so is a priori' and 'It is a priori that so and so' can be used to make *relative* ascriptions of a priority to propositions. For instance, we can conveniently take 'It is a priori that p' as ascribing a priority to the proposition that p *relative to the sentence* '*p*',[19] unless indicated otherwise. (Indeed, even Michael's assertion that he knows a priori that ## can be fully appreciated only if it is taken as involving implicit relativization. For Michael seems to think that one moral of the '##' example is precisely that *relative* a priority is called for under a certain conception of propositions.[20]) Hence we need not read the claim that I know that ## a priori or the claim that I know that Hesperus is Phosphorus a posteriori as entailing that I know that Mary was born in Seattle a priori or that I know that Hesperus is Hesperus (only) a posteriori. Taken as involving *implicit* relativization, these claims do not have such consequences. For knowing a priori that Mary was born in Seattle with respect to '##' does not entail knowing a priori that Mary was born in Seattle with respect to 'Mary was born in Seattle'.

4. To provide a yet wider perspective from which to appreciate that, the issue of the necessary a posteriori aside, a relative notion of an a priori proposition is genuinely needed in a philosophical logic informed by the direct reference theory, we can briefly consider singular propositions. The theory of singular propositions is widely recommended as the fine-grained account of propositions most congenial to the direct reference theory, the theory of reference behind Kripkean a posteriori necessity. Theorists of singular propositions claim that the contribution made by an ordinary name, or indexical (with respect to a certain context), to the proposition expressed by a sentence is simply the referent of the term. So, according to the theory, we can think of the proposition expressed by 'Aristotle is fond of dogs' as something like the ordered pair P, <Aristotle, being fond of dogs>. Now since P contains the flesh and blood Aristotle, to evaluate the truth value of P in a counterfactual circumstance *c*, it is obvious which individual in *c* is the one we should look at: Aristotle himself, rather than the unique object in *c* that happens to have the properties specified by a certain individual concept. Rigid designation is thus secured. Kaplan puts this point picturesquely:

> If the individual is loaded into the proposition (to serve as the propositional component) before the proposition begins its round-the-worlds journey, it is hardly surprising that the proposition manages to find that same individual at all of its stops, even those in which the individual had no prior, native presence. The proposition conducted no research for a native who meets propositional specifications; it simply 'discovered' what it had carried in. In this way we achieve rigid designation. (Kaplan 1989: 569)

Of course, a general proposition, for instance <C, the property F>, where C is an individual concept of some *individual essence*, whatever this may mean, will also manage to find the same individual at all its stops (or more accurately, all its stops where there exists an object having that essence), but the rigidity in this case is achieved

[19] I am ignoring the 'use/mention' convention in my use of " 'p' " here.
[20] See Michael (1998): 124.

by conducting 'research for a native who meets propositional specifications' (though such research 'happens' to find the same object, if it finds one, at each stop). Thus any term *t* expressing the concept C is semantically distinct from a name or any directly referential term: *t* does not contribute its referent to the proposition. Hence, when we think of direct reference in terms of the notion of a singular proposition, we have a clear account of how the rigidity of, say, names, is grounded on direct referentiality. The theory of singular propositions, we can thus say, provides a transparent way of explaining the deep structure of rigid designation of directly referential terms and stating the truth conditions of sentences containing such terms.

Most ordinary names are, according to the direct reference theory, genuine naming devices. Accordingly, a person, a tree, or a copy of *Word and Object* may literally be a part of a singular proposition. Just as there are different ways to represent a person, a singular proposition can be apprehended, or presented, or grasped, in different ways or in different guises. As a matter of fact, this general notion of, in Salmon's words, a 'guise' in which a proposition is apprehended is widely employed by proponents of singular propositions. As far as I know, in nearly all recent accounts of propositional attitudes proposed by those who espouse singular propositions, or who hold a view akin to Russellianism, there is some notion of a *mode of access*—a 'mediator' by means of which one is given access to a proposition, be it a 'guise' (Salmon 1986), a 'role' (Perry 1977), the 'content of cognitive states' (Fitch 1987), a 'sentence' (Soames 1989, Richard 1990), a 'character' (Kaplan 1989), or a 'nonlinguistic mode of presentation' (Geirsson 1994).

As I see it, this wide acceptance of a general notion of a mode of access by theorists of direct reference and theorists of singular propositions calls for a new, *relative* construal of the a priori. The notion of a guise or mode of presentation puts propositions *beyond the direct apprehension* of the knower. Thus in general a proposition can no longer be regarded as knowable in a direct, absolute sense, as it can and should be in the orthodox account of propositions.[21] For instance, the proposition *A* (i.e., the triple <identity, Venus, Venus>) can be said to be a priori only in some appropriate guise, say (the linguistic guise of) 'Hesperus is Hesperus'. The same proposition can at the same time be said to be a posteriori in another guise, say, 'Hesperus Phosphorus'.

In the kind of two-dimensional account suggested by Chalmers, Jackson, Stalnaker, or Wong, in which propositions are construed as sets of possible worlds or functions, the problem of guise may seem insignificant. However, this does not mean that attributing a priority directly to propositions or talking about a priori propositions in a non-relative way is not problematic for two-dimensionalists, as consideration of the dual-proposition objection reveals. The following parallel may help underscore this point. A proposition represented as a row in a certain two-dimensional matrix can also figure in other matrices. That is, just as a singular proposition can be apprehended in different guises, a proposition can be associated with different propositional concepts and thus with different two-dimensional

[21] I discuss the orthodox account of propositions in Wong (1991).

propositions. The insight behind the two-dimensional account is to explain the epistemic status of some necessary truths by appealing to this *multiple* association of a proposition with two-dimensional constructs. But this explanatory task will founder if we fragment a necessary a posteriori truth into two, for then we will fall into the trap of the dual-proposition problem. This fragmentation is unavoidable so long as we hold onto the absolute view. Hence the correct way to exploit the multiple association of a proposition with different two-dimensional constructs, I suggest, is to abandon the absolute view and to relativize propositional a priority and a posteriority to these constructs through sentences or other modes of access. The relative account is a natural step to take for both singular proposition theorists and two-dimensionalists, and it can easily be seen as such once we extract ourselves from the grip of the traditional, absolute view of the a priori and the a posteriori. A new paradigm of one thing often calls for a new paradigm of another. This, I think, is the case with the (already not so new) theory of direct reference and the relative account of the a priori and a posteriori.

References

Åqvist, L. (1973). 'Modal Logic with Subjunctive Conditionals and Dispositional Predicates', *Journal of Philosophical Logic* 2, 1–76.

Bar-Hillel, Yehoshua (1954). 'Indexical Expressions', *Aspects of Language*, Jerusalem: The Magnes Press, 1970, 69–88.

Chalmers, David J. (1996). *The Conscious Mind*, Oxford: Oxford University Press.

——(2004). 'Epistemic Two-Dimensional Semantics', *Philosophical Studies* 118, 153–226.

——(Forthcoming). 'The Nature of Epistemic Space', <consc.net/papers/espace.html>

Colterjohn, J. and MacIntosh, D. (1987). 'Gerald Vision and Indexicals', *Analysis* 47, 58–60.

Davies, Martin and Humberstone, I. L. (1980). Two Notions of Necessity', *Philosophical Studies* 38, 1–30.

Donnellan, K. S. (1983). 'Kripke and Putnam on Natural Kind Terms', in C. Ginet and S. Shoemaker (eds.), *Knowledge and Mind*, Oxford: Oxford University Press, 84–104.

Fitch, G. W. (1987). *Naming and Believing*, Dordrecht: D. Reidel.

Geirsson, Heimir (1994). 'Necessity, Apriority, and True Identity Statements', *Erkenntnis* 40, 227–42.

Hansson, Bengt (1974). 'A Program for Pragmatics', in S. Stenlund (ed.), *Logical Theory and Semantic Analysis*, Dordrecht: D. Reidel, 163–74.

Jackson, Frank (1998). *From Metaphysics to Ethics*, Oxford: Oxford University Press.

Kamp, J. A. W. (1971). 'Formal Properties of "Now" ', *Theoria* 37, 227–73.

Kaplan, David (1979). 'On the Logic of Demonstratives', in P. A. French *et al.* (eds.), *Contemporary Perspectives in the Philosophy of Language*, Minneapolis: University of Minnesota Press, 401–12.

——(1989). 'Demonstratives: an Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals' (with 'Afterthoughts'), in J. Almog *et al.* (eds.), *Themes From Kaplan*, Oxford: Oxford University Press, 481–563 (565–614).

Kripke, Saul (1972). 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: D. Reidel, 253–355. Also published as a book, *Naming and Necessity*, Oxford: Basil Blackwell, 1980. All page references given in this article are to the book.

Kripke, Saul (1979). 'A Puzzle About Belief', in A. Margalit (ed.), *Meaning and Use*, Dordrecht: D. Reidel, 239–83.
Lewis, David (1972). 'General Semantics', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: D. Reidel, 169–218.
____ (1981). 'Index, Context, and Content', in S. Kanger and S. Ohmann (eds.), *Philosophy and Grammar*, Dordrecht: D. Reidel, 79–100.
Michael, Michaelis (1998). 'Tichy on Kripke on A Posteriori Necessities', *Philosophical Studies* 92, 113–26.
Montague, Richard (1970). 'Pragmatics and Intensional Logic', *Synthese* 22, 68–94.
Perry, John (1977). 'Frege on Demonstratives', *Philosophical Review* 86, 474–97.
Plantinga, Alvin (1974). *The Nature of Necessity*, Oxford: Oxford University Press.
Quine, W. V. (1969). 'Proposition Objects', in *Ontological Relativity and Other Essays*, New York: Columbia University Press.
Richard, Mark (1990). *Propositional Attitudes*, Cambridge: Cambridge University Press.
Salmon, Nathan (1986). *Frege's Puzzle*, Cambridge, Mass.: MIT Press.
____ (1991). 'How Not to Become a Millian Heir', *Philosophical Studies* 62, 165–77.
____ (1993). 'Relative and Absolute Apriority', *Philosophical Studies* 69, 83–100.
Simpson, Paul (1987). 'Here and Now', *Analysis* 47, 61–62.
Soames, Scott (1989). 'Semantics and Semantic Competence', in J. E. Tomberlin (ed.), *Philosophical Perspective, Vol. 3*, Atascadero, Calif.: Ridgeview, 575–96.
Stalnaker, R. (1972). 'Pragmatics', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: D. Reidel, 380–97.
____ (1978). 'Assertion', in P. Cole (ed.), *Syntax and Semantics Vol. 9: Pragmatics*, New York: Academic, 315–32.
____ (1980). 'Logical Semiotic', in E. Agazzi (ed.), *Modern Logic—A Survey*, Dordrecht: D. Reidel, 439–56.
____ (1981). 'Indexical Belief', *Synthese* 49, 129–51.
____ (1999). *Context and Content*, Oxford: Oxford University Press.
____ (2001). 'Metaphysics and Conceptual Necessity', presented at II Barcelona Workshop: Two-Dimensionalism, Barcelona.
Tichy, Pavel (1983). 'Kripke on Necessary *A Posteriori*', *Philosophical Studies* 43, 225–41.
Vision, Gerald (1985). 'I am Here Now', *Analysis* 45, 189–99.
____ (1987). 'Antiphon', *Analysis* 47, 124–8.
Wong, Kai-Yee (1990). 'Reference, Context, and Propositions', Ph.D. thesis, Australian National University.
____ (1991). 'A Priority and Ways of Grasping a Proposition', *Philosophical Studies* 62, 151–64.
____ (1996a). 'Sentence-Relativity and the Necessary A Posteriori', *Philosophical Studies* 83, 53–91.
____ (1996b). 'Singular Propositions and the A Priori', *Journal of Philosophical Research* 21, 107–16.

15

No Fool's Cold: Notes on Illusions of Possibility

Stephen Yablo

A lot of philosophers are *pessimistic* about conceivability evidence. They think it does not prove, or even go very far towards justifying, interesting modal conclusions. A number of other philosophers are *optimistic*; they think it does justify, and perhaps even establish beyond a reasonable doubt, that lots of interesting things are possible. Nothing very surprising there. What is slightly surprising is that both groups can claim to find support for their attitude in the work of Saul Kripke.

Pessimists say: Kripke shows that conceivability evidence is highly and systematically *fallible*. Very often E seems possible, when as a matter of fact, E-worlds cannot be. So it is, for instance, with the seeming possibility of water in the absence of hydrogen, or of Hesperus distinct from Phosphorus, or of this table turning out to be made of ice. Let the pessimistic thesis be

(P) oftentimes E seems possible when it is not, so conceivability evidence is not to be trusted.

Optimists reply: yes, Kripke finds conceivability evidence to be fallible, but that is only half of the story. The rest of the story is that *the failures always take a certain form*. A thinker who (mistakenly) conceives E as possible is correctly registering the possibility of *something*, and mistaking the possibility of *that* for the possibility of E. There are *il*lusions of possibility, if you like, but no *de*lusions or hallucinations. Let the optimistic thesis be

(O) Carefully handled, conceivability evidence can be trusted, for if impossible E seems possible, then something else F is possible, such that we mistake the possibility of F for that of E.

The optimistic thesis (O) represents conceivability evidence as in a sense *in*fallible. If (O) is correct, then that E seems possible, while it may not establish that E is possible,

This paper was presented at the UNC Greensboro conference on imagination and possibility, with comments by Keith Simmons. Thanks to Keith for exposing various gaps in the argument, not all of which I have been able to deal with here. Thanks to Kit Fine, Tamar Szabó Gendler, Janine Jones, and Saul Kripke for discussion at the conference, and to David Chalmers and Tyler Doggett for extremely helpful written comments provided more recently.

does succeed in establishing the disjunctive conclusion that either E is possible or F is. And indeed in certain cases we can get all the way to the first disjunct, because F is tantamount to E or entails E. This, the optimist continues, is the situation we encounter in the last few pages of *Naming and Necessity*, where Kripke argues against the identity theory of mind. It seems possible that pain is not c-fiber firings, and the F that supposedly snookers us into thinking E possible is tantamount to that original E. (I will be questioning that argument in due course.)

It seems likely that both groups are overinterpreting Kripke. Certainly Kripke is not a pessimist, because he closes the book with a positive argument of the sort that pessimists are bound to find fault with. And although this is not as clear, he seems to stop short of outright optimism too. He says (in "Identity and Necessity") that "the only model *I can think of* for what the illusion might be. . . does not work in this case" (1977, 101, emphasis added). Others are welcome to argue in favor of some other model that does not require a genuinely possible F. Kripke is skeptical, to be sure: "it would have to be a deeper and subtler argument than I can fathom and subtler than ever appeared in any materialist literature that I have read" (1977, 101). But although Kripke has his doubts about the availability of an alternative model, he does not entirely rule it out. (One is reminded of Carnap's position in "Empiricism, Semantics, and Ontology": I can't make sense of the question of realism my way; maybe others can find a different way, but it won't be easy.)

So the door is open, technically anyway, to "a deeper and subtler argument" aimed at establishing that *some* seeming possibilities do not reflect *any* sort of genuine possibility. Whether this deeper and subtler argument can be given has not been terribly much explored.

****

One idea sometimes encountered is that there are differences in how pains and c-fiber firings are entertained in thought that *all by themselves* explain why each would seem possible without the other. Thomas Nagel's version of this idea is that C-fiber firings are imagined perceptually—"we put ourselves in a conscious state resembling the state we would be in if we perceived it"—while pain is imagined sympathetically—"we put ourselves in a conscious state resembling the thing itself" (1974, note 11). He maintains that:

> the relation between them will appear contingent, even if it is necessary, because of the independence of the disparate types of imagination. (1974, note 11).

Chris Hill says in a similar vein that the relation appears contingent because our concept of c-fiber firings is theoretical while our concept of pain is phenomenological. Between concepts like that "there are no substantive a priori ties," and the absence of such ties allows us to "use the concepts to conceive coherently of situations . . . in which there are particulars that fall under one of the concepts but do not fall under the other" (1997, 75).

This sort of approach is in one way too broad and in another too narrow. It is too broad in that it threatens to undermine conceivability arguments that most of us find attractive. It certainly *seems* to me that my dog Ruby could have been in severe

pain right now; that's what you normally get for harassing a porcupine. But then so it would, according to Nagel, what with Ruby being imagined perceptually and the pain sympathetically.

I agree that the appearance here should not be taken seriously, if it arises in the way Nagel says. That we do take it seriously suggests that the explanation may not be quite so simple. And indeed there are independent reasons to think matters are not so simple. If appearances of contingency resulted just from "disparate types of imagination," then one would expect more to seem possible than in fact does. After all, it is not just the dog that is imagined perceptually but everyday objects in general. Consider the rock that Ruby is perched on. All the Nagelian conditions are in place, yet it does *not* seem that the rock could have been in pain right now. It takes more to tempt us into an illusion of possibility than Nagel supposes.

What about Hill's version of the idea? It seems to me, as I consider this cup of vinegar, that a cup of H_2O could look just the same. But then so it would, on Hill's view, for *looking the same* is a phenomenological concept, while our concept of H_2O is theoretical. Once again, though, this cannot be all there is to it, for there are cases where Hill's conditions are met and the appearance of contingency is lacking. A cup of molten lead does *not* present itself as capable of looking like this.[1]

How is the Nagel-type approach too narrow? By focusing so intently on subjective versus objective, it just reinforces the impression that Kripke is trying to create, namely that any response to his argument is going to require some kind of special pleading on behalf of the mental. I cannot rule it out, of course, that the proper response *does* require special pleading. But it would be better if we could identify a general constraint on modal illusions that is independently motivated and that just happens to deliver the desired results when applied to the intuitions supporting mental/physical dualism.

****

I want to explore some of these issues by looking at the role of *actuality* in modal judgments. Actuality comes in under two separate headings. On the one hand it can figure in the *content* of a modal judgment. The thing that seems possible—the condition that seems like it could have obtained—can have the notion of actuality in it. This is in fact quite common. One says, for instance, "this lemonade is cold but it could have been colder."[2] Colder than what? Colder than it actually is, of course. If C is the "how cold was it?" parameter, then our judgment is roughly this

seems $\Diamond$ (C exceeds $C_@$)

Or perhaps we are doing a puzzle where five irregularly shaped pieces of plastic have to be rearranged into a square. We look the pieces over and it strikes us that the thing

[1] Tyler Doggett and Daniel Stoljar point out that the Nagel worry also pulls the rug out from under standard objections to behaviorism and functionalism. Given any behavioral property B, we can imagine being in pain without exhibiting B and vice versa. Perhaps the appearance of contingency here is due just to the fact that pain is imagined sympathetically and B perceptually.

[2] Could have been colder as a liquid, I mean. Assume for the sake of the example that so-called frozen lemonade is not really lemonade.

can be done. What seems possible, however, is not that the pieces can be made to form a square *after being melted down and recast as rectangles*; it's that they can be made to form a square with their actual shapes and sizes held fixed. If the shape and size of piece X is S(X), then our judgment is

seems ◊ (the Xs form a square & ∀X (S(X) = $S_{@}$(X)))

A remark attributed to Richard Taylor gives us a third example. "Why are people so sure they could have acted otherwise?" he asks. "After all, nobody ever has." One reason we think this is that it very much *seems* as though we could have acted otherwise:

seems ◊ (my action was of a type T incompatible with the type $T_{@}$ of the action I really did perform)

To have a schema for judgments of this kind, what seems possible is that a certain parameter P should have taken a value so and so related to the value it actually takes:

seems ◊ (. . . & P is so and so related to $P_{@}$ & . . .)

That is the first way actuality can come in. It leads pretty directly to a second way. Whether or not it seems possible for some parameter to assume a value so and so related to its actual value is not independent of what we know, or think we know, about what the actual value in fact is, or indeed of other information we possess about actuality. It would not have seemed possible for the pieces to be rigidly rearranged into a pentagon if we had believed each piece to be square, or round. It would not have seemed possible for the lemonade to be colder if it was believed to be at zero degrees already. It might not have seemed possible for us to act otherwise were we convinced that Frankfurt's nefarious neurologist (made omnipotent if necessary) stood ready to reprogram our brains if we tried.

There is a temptation, perhaps, to treat this as just more content. But the temptation should be resisted, because it imports more into the content than belongs there. Our judgment is not

seems ◊ (this lemonade is colder than N° C).

After all, we may have little positive idea what temperature the lemonade is in degrees centigrade. What seems possible is that the lemonade should be colder than it is, and why it seems possible has to do with the lemonade's felt temperature.[3]

If our sense of the temperature doesn't figure in content, though, what role *does* it play? It plays what might be called a *presuppositional* role. The judgment is conditioned on our temperature experience's not being too misleading. One thinks, "unless I am very much misled about how cold this liquid is, it could have been colder." Besides appearing in the *content* of a modal judgment, then, actuality can figure in the *background* to the judgment, that is, the beliefs or presuppositions that allow the seemingly possible thing to seem possible.

****

[3] Specifically, with its feeling warmer than lemonade on the verge of freezing feels.

Back now to the main issue. The optimist says that whenever there is the illusion that *E* is possible, there is a related hypothesis *F* that really is possible. For instance, it seems that Hesperus could have been distinct from Phosphorus because there really could have been two planets there, one responsible for Hesperus-appearances and the other for the appearances we enjoy of Phosphorus. I have said a little about *E*, the content of the (perhaps mistaken) intuition, but nothing about *F*, the hypothesis that is supposed to really be possible.

Kripke does not even pretend to give us a general strategy for recovering *F*—what I will call *the underlying possibility*—from *E*. What he does do is, first, sketch lots of highly convincing examples; second, suggest that at least some of the time, it is good enough to replace names in *E* with corresponding reference-fixing descriptions; and third, characterize *F* as the "appropriate corresponding qualitative contingent statement." He explicitly refrains, though, from giving a "general paradigm" for the construction of the proposition whose possibility fools us into thinking *E* possible.

A number of other writers have been bolder. Some say that there is the illusion that *E* is possible because the sentence "*E*" could (with its "meaning" in some sense of that word held fixed) have expressed a true proposition, albeit not the proposition it expresses in fact. So,

(a) it could have happened that "*E*" expressed a true proposition.

I myself once conjectured that *E* seems possible because we could have thought something true with the thought (the internal mental act) whose content in this world is *E*. So,

(b) it could have happened that thinking the *E* way was thinking truly.

The best-known suggestion along these lines is that *E* seems possible because there are worlds such that if (contrary to what we perhaps suppose), they are actual, then *E*. So a third hypothesis is that

(c) things could have been a way such that, if they actually *are* that way, then *E*.

All these proposals are variations on the theme of *E* seeming possible because what it says is correct, if a certain not-impossible world is actual. Nothing important is lost if we ignore any differences and speak simply of the *if-actually* account of illusions of possibility.

The if-actually account works extremely well in some cases. The reason it seems possible that the table should turn out to be made of ice is that there are worlds with the property that if they are actual, then it *is* made of ice. The reason it seems possible that Hesperus should have been other than Phosphorus is that there are worlds with the property that if they are actual, it really is other than Phosphorus. It turns out, though, that the account cannot deal correctly with actuality-based modal contents. I will build up to this slowly.

Ivory-billed woodpeckers had been thought extinct; recently, though, a man named David Kullivan reported spotting a pair of them. I happen to believe this report, but not everyone does. Knowing that his word would be doubted, Kullivan was tempted (let us say for purposes of the example) to shoot one of the woodpeckers

and bring its body back as proof. According to me, believing as I do that ivory-billed woodpeckers exist, had Kullivan shot one, there would have been fewer ivory-billed woodpeckers than there are. To me, then

seems ◊ (there are fewer ivory-billed woodpeckers than actually).

Now suppose that I am wrong and there are no ivory-billed woodpeckers. Then I am under an illusion of possibility; a smaller number seems possible, but there cannot be fewer than none. What explains my illusion? The story would have to be that this seems possible because there is a world such that if it is actual, then there *are* fewer ivory-billed woodpeckers than there actually are. And that makes no sense.

Of course, there is no peculiarly *modal* illusion here; where I go wrong is in believing in ivory-billed woodpeckers in the first place. But consider a second example. It seems possible that Hesperus could have turned out to be distinct from Phosphorus. It seems, for instance, that Phosphorus could have turned out to be Mars rather than Venus. Another thing that seems possible is for Phosphorus to have turned out to be *Xorg*, a solar planet over and above the planets that exist in fact. It seems possible, then, that there should have been more planets than actually: all the actual ones, including Hesperus, and then in addition Phosphorus = Xorg.

seems ◊ (there are extra planets; Hesperus is Venus but Phosphorus is new).

The story would have to be that this seems possible because if we are wrong and the morning-visible planet is "new," then there really *are* more planets than actually. And that clearly cannot be right. Again, it strikes us that gold could have turned out to have a different chemical makeup. The illusion that gold could have failed to be the 79th element *can* be explained, notice. But I may not know that gold is any kind of element; my thought is just that it did not *have* to turn out with that chemical makeup, whatever its makeup in fact is. This illusion cannot be explained on the if-actually model, for we would need a world such that gold has a different makeup than it actually does on the supposition that this world is actual.

****

So the if-actually account cannot explain certain illusions of possibility, those in which the hypothesis that seems possible involves a contrast or comparison with actuality.[4] Why should we bother about this? The reason for bothering is that it tells

[4] One natural idea about actuality-involving illusions (suggested independently by Robert Stalnaker and David Chalmers) is this: they are to be explained by saying there is a world *w* such that if *w* is actual, then the actuality-involving proposition is *possible*. It seems possible for there to have been fewer ivory-billed woodpeckers because this really is possible on the hypothesis that Kullivan's story is true. But the intuition that Hesperus could have been an additional planet is not based in any factual misinformation of the sort we might try to correct by treating *w* as actual. The feeling is not that assuming Phosphorus is other than Hesperus, it could have been Xorg. The feeling is that Phosphorus, although (it turns out) identical to Hesperus, could have been distinct from it in a way that bumped up the number of planets.

us something about how people are thinking of the modal illusion problem. The if-actually account is exceedingly popular. (I stress that Kripke does not endorse it.) Why, if there is a class of illusions it does not address? *It must be that this class of illusions has not been much on people's minds*. People have been assuming, implicitly anyway, that the contents of error-prone modal judgments are *actuality-neutral* in the sense, roughly, that facts about which world is actual are irrelevant what the judged hypothesis says. Perhaps to be safer I should just say that there has been a tendency to downplay or underestimate the actuality-based aspects of these contents, and to play up or overestimate their actuality-neutral aspects.

One sort of problem this bias in favor of neutrality leads to has already been seen. But the problem that interests me is not that certain actuality-based illusions will prove difficult to explain, but that certain such illusions will be "explained" too easily. This is how it would happen:

(1) What seems possible is a hypothesis E that is actuality-based.

(2) An actuality-neutral (or more neutral) hypothesis E' is covertly substituted.

(3) One explains the illusion that $\Diamond E'$ as a subtle misreading of $\Diamond F'$.

(4) It would take a very much *grosser* misreading of $\Diamond F'$ to fall under the illusion that $\Diamond E$.

(5) One thinks the E illusion has been explained when really it has not.

I will give examples in a minute. But first let me link the worry up with what I take to be an important feature of Kripke's procedure.

Kripke does not just want to show how someone *could* fall under the misimpression that, say, Hesperus could have failed to be Phosphorus, by misinterpreting what was in fact a different possibility. That would be easy, since a sufficiently confused person could presumably misinterpret anything as anything. He wants to show that we plausibly *do* fall under the modal misimpression by misinterpreting a different possibility. It is not just that an intuition of E's possibility *could*, but that our intuition of its possibility plausibly *is*, based on the mistaking of one possibility for another.

An example of someone who seems to underestimate the aspiration here is Michael Della Rocca in "Essentialism and Essentialists" (*Journal of Philosophy* 1996). Say that Lumpl is the lump of clay composing the statue Goliath. It seems possible that Lumpl could have failed to be Goliath, or any other statue; it seems possible, indeed, that Lumpl could have existed *in the complete absence of statues*.

(a) seems $\Diamond$ (Lumpl exists without any statues)

Della Rocca maintains that this intuition is (or might be for all Kripke has to say about it) explained by the possibility that *a lump of clay handled by artisan A at time T* should have lacked all these properties.

(b) really $\Diamond$ (a lump handled by A at T exists without any statues).

I suppose that (b) *might* perhaps explain the illusion of someone for whom the reference of "Lumpl" was fixed by "the lump of clay handled by *A* at *T*." But "Lumpl" in *our* mouths has its reference fixed by "the lump composing the statue Goliath." (That is how I introduced the term above, and that is the usual way of introducing it.) So, the genuine possibility needed to explain away *our* intuition is

(c) really ◇ (a lump composing the statue Goliath exists without any statues)

But there is no such possibility as (c); it cannot happen that a lump both composes a certain statue and fails to coexist with any statues. The scenario that (c) calls possible, and whose possibility would be needed to explain away the intuition that Lumpl could exist without statues, makes no sense.

I seriously doubt, then, whether our *actual* intuition of Lumpl without statues can be defeated as easily as Della Rocca suggests.[5] The only real possibility in the neighborhood is the one recorded in (b). And there is no way on earth that we are misinterpreting *that* as the possibility of Lumpl without any statues. The proof that (b) does not explain (a) is just that stare at (b) as long as you like, one cannot imagine being so confused as to have been fooled by it into supposing that (a). One is not at all tempted to say: oh, I see, once you point out the difference, it's because *this* really is possible that I supposed *that* to be possible.[6]

The kind of principle I am relying on here is familiar from psychoanalysis. Here is what in my brief (well, . . .) experience psychoanalysts tell you. "You are under the impression that nobody loves you. I submit that this is an illusion. A cruder sort of doctor might say, here is how the illusion arises, take my word for it. But I would never dream of asking you to take my word for it. No, the test of my explanation is whether you can be brought to accept the explanation, and to accept that your judgment is to that extent unsupported." The analogy is good enough that I will speak of the

> *Psychoanalytic Standard* Assuming the conceiver is not too self-deceived or resistant, ◇*F* explains *E*'s seeming possibility only if he/she does or would accept it as an explanation, and accept that his/her intuition testifies at best to *F*'s possibility, not *E*'s.

[5] Della Rocca brushes up against this problem in a footnote. "One might, perhaps, see some other property as the property in terms of which Lumpl is identified. Even if some other property is the identifying property, the argument that I am about to give would not be affected because I shall show that *any* property that might plausibly be seen as the property in terms of which Lumpl is identified would be a property that allows a Kripkean reconstrual of our intuition of contingency in this case to go forward" (197). I do not see that he ever shows this. What he does say is that "Lumpl seems to be identified in terms of the designation, 'lump formed by, etc.', or some similar designator. Any such designator would allow the reconstrual to go through" (197–8). This is false, unless "similar" means "designator H such that there could be an H without Goliath existing." The designator "clay composing Goliath" is an obvious counterexample.

[6] Della Rocca agrees that the (b) possibility is not *judged* explanatory. He thinks, however, that any attempt to justify this judgment winds up begging the question at issue: which modal intuitions are windows on possibility and which are illusions of possibility?

This is a high standard, but what makes Kripke's approach so convincing is that this is the standard he tries to meet, and mostly *does* meet. Philosophers have been telling us for centuries that this or that common impression is false; and we have for centuries been shrugging them off. What makes Kripke special is that he gets you to *agree* that you are making the mistake he describes.

****

I said that Kripke "mostly" meets the psychoanalytic standard. This is because I think that with at least some of the illusions he discusses, the standard is *not* met, and is perhaps unmeetable. Let me start with an example where a psychoanalytically acceptable explanation *can* be given. I will then argue that a crucial feature of the example goes missing in Kripke's treatment of certain other examples.

Kripke says, ". . . though we can imagine making a table out of another block of wood or even from ice, *identical in appearance to this*, and though we could have put it in this very position in the room, it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, resembling this one in all external details, made of another block of wood, or even of ice" (1980, 114, emphasis added).

Imagine someone, call them Schmipke, expressing puzzlement about Kripke's procedure: "Hasn't Kripke gone to a lot of unnecessary trouble here? Why does he impose this condition of *identical in appearance* with the actual table? 'Identical in appearance' suggests that the otherworldly table looks just like the real one *to us*: if both of them were sitting here side by side, we could not tell them apart. This is suggested as well by the language he uses in 'Identity and Necessity': "I could find out that an *ingenious trick* has been played on me and that, in fact, this lectern is made of ice" (1977, 88). The ice has to be 'cleverly hardened' in the shape of a table, and presumably painted too. Otherwise it would not be a spitting image of our actual table, as Kripke clearly intends. Is any of this really necessary? Why does Kripke ask *w* to satisfy the *actuality-based* condition that its table looks or would look just the same to us? What is wrong with the *neutral* condition of, not identical *in* appearance, but simply: identical appearance*s*?"[7]

This seems a fair question, so let us try it. Until further notice, all we require from *w* is that there is an icy table there, and that the people looking at it (perhaps counterfactual versions of ourselves) have the same experiences qualitatively speaking as we do looking at our table. It is of course compatible with this that the tables look to *us* very different. But then our reason for thinking of the icy table in *w* as "in disguise," cleverly tricked up to look like wood, no longer applies. Now that we have dropped the identical-in-appearance requirement, the icy table can be made any number of ways. Let it be, say, a table-shaped, table-sized, but otherwise perfectly ordinary frosty white block of ice. Of course, it needs to be added that the observers in *w* are spectrum-inverted with respect to observers here, so that the qualitative

[7] Or, if that is not neutral enough, let the condition be not that observers in *w* enjoy qualitatively identical appearances, but that they enjoy qualitative appearances PQR. I will ignore this complication.

appearances they enjoy in front of a frosty white object are just like the ones we enjoy when looking at an otherwise similar brown object. But if both of those changes are made at once, then the experience of observers there looking at their table is just like the experience we enjoy looking at ours.[8]

Note that there is *some* slight support for Schmipke's position in the text. Kripke says that what the icy table intuition comes to is that "I (or some conscious being) could have been qualitatively in the same epistemic situation that in fact obtains, etc." *He does not say the conscious being has to resemble me in any important respect*. The counterfactual being's brain might be wired so that it is in the same qualitative state standing in front of an icy table as I am standing in front of a wooden one. So, contrary to what we said above, it could be that Kripke is imposing only the neutral condition of *icy table*, *appearances XYZ*.

The question is, does the revised explanation meet the psychoanalytic standard? Does it explain our illusion that *this table could have turned out to be made of ice*, to point out that had our brains been different, a regular icy table would have caused in us the same qualitative state that a wooden table does cause in us? I tend to think it does not. Because what seems possible is that this table *with relevant perceptible properties held fixed* could have turned out to be ice. No one is going to be tempted into thinking *that* possible by reflection on the possibility that we see a regular icy table as brown, because in that scenario the perceptible properties *change*. The color of the table goes from brown to white.[9]

It may help to consider an analogy. Say that I am under the impression that that animal there [pointing] is a zebra, when really it is a horse. Dretske's explanation is this: "The horse is painted to look just like a zebra. When two things look just the same, the one is easily mistaken for the other. It makes sense then that you would take this horse for a zebra." That corresponds to the Kripkean explanation of the "could have turned out to be ice" illusion. Because the table's appearance is indistinguishable from that of disguised ice, one naturally concludes that it could *be*, or have been, disguised ice.

Imagine now a second, Schmipkean explanation of my zebra illusion. "The horse is not painted at all. And you're enjoying ordinary horsy phenomenology. But there is this guy counter-Steve, a counterfactual variant of yourself, who has zebraish phenomenology when looking at a horse, and horsy phenomenology when looking at a zebra. Because your phenomenology is indistinguishable from that of counter-Steve looking at a zebra, it makes sense that you would take this horse for a zebra." That corresponds to the Schmipkean explanation of the "could have turned out to be ice" illusion. Because my actual table phenomenology is indistinguishable from my alter ego's ice phenomenology, I am led to suppose that this table could be, or have been, a regular old hunk of ice.

[8] Schmipke concedes the possibility of spectrum inversion.

[9] A property is perceptible iff when an object perceptually appears to have it and does not, we have misperceived. Not all properties figuring in the content of a perceptual state are perceptible in this sense. Our experience may represent the table as wooden, but it is not as if our eyes are playing tricks on us if it is well-disguised ice.

Is it just me, or does the first pair of explanations work better than the second? "I am liable to confuse A with B because they look the same to me" sounds quite plausible. If things look the same, then one is indeed liable to confuse them. "I am liable to confuse A with B because the same looks result if it is me looking at A or counter-Steve looking at B." *There is no chance at all* that I am confusing myself with counter-Steve, even if his phenomenology is just the same. Counter-Steve is by definition a person who sees things differently than I do. (One might as well worry that our planet has all along been Twin-Earth, making water not H_2O but XYZ.)

So we have the following principle: to explain why *this*, understood to present like so, seems like it could turn out to be Q, one needs a possible scenario in which something *superficially indistinguishable* from it does turn out to be Q. The counterfactual thing has to look the same, not to the counterfactual folks, but to us. I will call that a *facsimile* of the actual thing. And I will refer to the principle as the facsimile or fool's gold principle.

****

Kripke gives two models for the explaining-away of the intuition that A could be Q. First is the reference-fixer model:

(RF) it seems possible for A to be Q because it really is possible that the so and so is Q, where "the so and so" is a descriptive condition fixing "A"'s reference.

Then there is the epistemic counterpart model:

(EC) it seems possible for A to be Q because it really is possible for A* to be Q, where A* is a facsimile of A.

The epistemic counterpart model might seem the more accommodating of the two, because it does not require anything in the way of reference-fixing descriptions. But there is a respect in which the reference-fixing model is more accommodating and indeed *too* accommodating.

The epistemic counterpart model requires an A* indiscernible in relevant respects from A, what we have called a facsimile of A. Can this requirement be enforced by asking A* to satisfy some carefully constructed reference-fixing description D? It is not at all obvious that a suitable D can be found. One obvious possibility is "the thing that puts me into qualitative state 279." The picture this gives is:

me-in-@) $QS_{279} \longrightarrow A$
me-in-*w*) $QS_{279} \longrightarrow A^*$

Here we have dissimilar observers in distinct worlds confronting two (perhaps readily distinguishable) objects and reacting the same way. (EC) by contrast envisages a single observer confronting two objects to which she responds identically:

$$\text{me-in-@) } QS_{279} \begin{array}{l} \longrightarrow A \\ \longrightarrow A^* \end{array}$$

Perhaps we can arrange for the second picture by letting D be the "the thing that puts me *as I actually am* into qualitative state 279." But this forgets that "the thing that actually puts me in state 279" stands in counter-Steve's mouth for A*. We are left again with the first picture.

One could try to force the second picture by letting D be "the thing that *in* α puts me into state 279," where α is a stable designator of actuality; it picks out our world @ no matter in which actual or counterfactual context it is uttered. But the point of a reference-fixing description is that it is supposed to be a piece of language that directs us to the referent across a range of counterfactual situations. And the term "whatever in α puts me into state 279" is not even *understandable* in counterfactual situations. Had things been different, we would not have been thinking, "too bad things are so different here, how much better to live in a non-counterfactual world like α."

Two pictures have been sketched of how to explain away modal illusions. Which of the two is meant to apply in the case of the icy table? Passages like "I (*or some conscious observer*) could have been in qualitatively the same epistemic situation" (1980, 142, emphasis added) suggest the first picture. But there are also passages like this:

> . . . it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, *resembling this one in all external details*, made of another block of wood, or even of ice. (1980, 114, emphasis added)

"Resembling in all external details" means, I take it, that we would not notice if the one table were instantaneously substituted for the other. And that is the second picture. The reason this matters is, once again, that the first picture fails to explain the illusion. It defies credulity that my feeling that *this table* could have been made of ice is based on the fact that *my brain* could have been such that suitably carved ice elicited in me the present sort of appearances.

****

But let us not dwell too long on the icy table example, since Kripke uses it mainly for illustration. His real interest is in the kind of modal illusion that arises in science. Here is some heat; is it some type of molecular energy?[10] One has to conduct further tests, and like any tests, they could come out either way. So there is the appearance that heat could turn out to be a certain type of molecular energy, and the appearance that it could turn out to be something else. The second appearance is an illusion. How does Kripke propose to account for it?

> the property by which we identify [heat] originally, that of producing such and such a sensation in us, is not a necessary property but a contingent one. This very phenomenon could have existed, but due to differences in our neural structures and so on, have failed to be felt as heat. (1980, 133)

It might be, for instance, that due to differences in our neural structures *high* mean molecular energy—henceforth HME—felt cold, and *low* mean molecular

[10] Like Kripke, I will run heat together with temperature.

energy—henceforth LME—felt hot. Does this explain in a psychoanalytically satisfying way our feeling that it could have been LME that was heat rather than HME? Does pointing to possible differences in our neural structures explain why this cold seems like it could have turned out to be HME?

Here is the worry. With the table, remember, what seemed possible is not only that ice could have paraded itself in front of *someone or other* who saw it as I see wood, but that there could have been ice that *I with my existing sensory faculties* would have seen as wood. To explain *that* seeming we needed a facsimile of the table—a spitting image of it—that was in fact ice. Likewise what seems possible in the case of LME is not just that it could have paraded itself in front of *someone or other* who felt it as hot, but that *I* with my existing neural structures could have found it to be hot. To explain that seeming, we need a counterfactual facsimile of heat that turns out on closer inspection to be LME. There should in other words be the possibility of LME-type fool's heat. Similarly, to explain the seeming possibility of cold turning out to be HME, we would need the possibility of fool's cold that was found by scientists to be HME.

Is there fool's heat of this type, or fool's cold? I do not see how there could be. It may be possible to slip a cleverly disguised icy table in for this wooden one with no change in visual appearance. *But it is not possible to slip cleverly disguised LME in for HME and have it feel just the same*. Having substituted low ME for high, there is no way to preserve the appearances but to postulate observers who react differently than ourselves to the same external phenomena. But then what we are getting is not really fool's heat but something more like *dunce's* heat. You would have to be pretty confused to see in the possibility of rewiring on *your* side the explanation of why a switcheroo seems possible *on the side of phenomenon you are sensing*. Whether fool's heat is absolutely impossible I don't know. But what does seem clearly impossible is for *LME to be fool's heat*, because it by hypothesis feels the opposite of hot; it feels cold.

****

Kripke is right, or anyway I am not disagreeing, when he says that "the property of producing such and such a sensation in us . . . is not a necessary property," because we could have been wired differently. LME could, it seems, have produced what we call sensations of cold. That is not what I am worried about. What worries me is that the property of interest is not that but *producing such and such a sensation in us as we are*. And this property is, I suspect, necessary. There would seem to be three factors in how an external phenomenon is disposed to feel: its condition, our condition, and the conditions of observation. If all these factors are held fixed, as the notion of fool's heat would seem to require, then it is hard to see how the sensory outcome can change.

Someone might say: that LME can't be fool's heat doesn't show that there can't be fool's heat at all. Surely there is *something* in some faraway world that although not HME feels or would feel hot to us as we are. Suppose that is so,[11] and call the something ABC ("alien basis caliente"). ABC is all you need to explain the illusion

[11] Kripke actually discusses something like this in *Naming and Necessity*. "Some people have been inclined to argue that although certainly we cannot say that sound waves 'would have been

that heat could have been other than HME in the approved Kripkean fashion, that is, in terms of a genuine underlying possibility.

But, granted that one *can* explain, or try to explain, the illusion in this way, would the explanation be correct? I am not sure that it would, for the following reason. Our feeling that heat could have turned out to be something else is *indifferent* to whether the something else is alien ABC or actual LME. It would be very surprising if the feeling had two radically different explanations depending on the precise form of the something else. The LME form of the illusion *cannot* be explained by pointing to a possible facsimile of heat that really is LME. (Whether LME can be fool's heat is a factual question and the answer is that it can be at best dunce's heat.) Therefore the ABC form of the illusion ought not to be explained with a possible facsimile either.

I have been arguing that *strong* epistemic counterparts, or facsimiles, are needed to explain illusions of possibility. However there are some illusions to which epistemic counterparts, strong or weak, might seem altogether irrelevant. It seems possible not only that heat could have failed to be HME, but also that HME could have failed to be heat. Kripke treats the latter illusion as reflecting the genuine possibility that HME might not have felt hot. Given that epistemic counterparts do not figure here at all, the insistence that any epistemic counterparts should be strong may seem to leave Kripke's explanation untouched.

Once again, I appeal to the principle that similar intuitions should receive similar explanations. Our intuition that HME could have turned out to be *something* other than heat differs only in specificity from the intuition that it could have turned out to be *cold*. Weak epistemic counterparts of cold are of no use in explaining the latter illusion; it does not matter what "those people" (the residents of *w*) think. But if otherworldly observers are irrelevant here, then they are irrelevant to the unspecific intuition as well.

The upshot is that if S is a sensed phenomenon like heat, and P is a physical phenomenon like LME, then otherworldly observers are no use in explaining *either* why S seems like it could have been other than P, or why P seems like it could have been other than S. Since, as we have seen, actual observers cannot explain these apparent contingencies either, it seems that there is no psychoanalytically satisfying explanation in Kripke for the appearance that S is only contingently related to P.

But, someone might say, this just shows we have been going about it the wrong way around. Rather than looking for a strong epistemic counterpart of *heat* that is LME, we should be looking for a strong epistemic counterpart of *me* to whom LME feels hot.

I do not deny that such a person is possible; the question is what he can do for us. It seems not an accident that the intuitions explained by facsimiles of the table are

heat' if they had been felt by the sensation which we feel when we feel heat, the situation is different with respect to a possible phenomenon, not present in the actual world, and distinct from molecular motion. Perhaps, it is suggested, there might be another form of heat other than 'our heat', which was not molecular motion; though no actual phenomenon other than molecular motion, such as sound, would qualify. Although I am disinclined to accept these views, they would make relatively little difference to the substance of the present lectures. Someone who is inclined to hold these views can simply replace the term . . . 'heat' with . . . 'our heat'. . . (p. 130, note 68).

intuitions about what is possible for *the table*. Likewise the intuitions explained by gold-facsimiles are intuitions about *gold*, for example, that it could have turned out to be iron pyrites. One would expect, then, that the intuitions explainable by reference to me-facsimiles are in the first instance intuitions about me. Am I the sort of person who has heat sensations in response to HME, or the sort of person to whom LME feels hot? There is the feeling (suppose for argument's sake that it is an illusion) that I could have been the second sort of person. How does this feeling arise? Well, a possible strong epistemic counterpart of mine *does* have heat sensations in response to LME.

But it is one thing to explain apparent de re possibilities for ourselves, another to explain apparent de re possibilities for heat. When we ask, "did heat have to be HME or could it have been LME?", and answer that it could have turned out either way, we are caught between two seeming possibilities *for heat*. The proof of this is that the seeming possibility of heat being LME does not depend in the least on there being Steve-like beings around to whom LME feels hot. (Perhaps heat's being LME creates conditions inhospitable to life.) The intuition that heat could have been LME *although there was no one around to realize it* cannot be explained by pointing to a possible me-facsimile reacting differently to LME, simply because it is stipulated in the intuition that no observers are present.

****

Here is the position so far. It is not hard to disguise a genuinely icy table so that it looks wooden. So if Kripke wants to explain the seeming possibility of this table A being made of ice, he has at his disposal a facsimile A* of the table that really is made of ice. Sometimes, though, the appearance is closer to the reality, and facsimiles of A are no more capable of possessing the seemingly possible property Q than A is itself. How the second sort of illusion arises is an interesting question, but a question for another paper.[12] The claim for now is just that we cannot explain the second sort of illusion by pointing to a world where an A-facsimile really is Q, because such a world is not possible.

Kripke says, "perhaps we can imagine that, by some miracle, sound waves somehow enabled some creature to see. I mean, they gave him visual impressions just as we have, maybe exactly the same color sense. We can also imagine the same creature to be completely *insensitive* to light (photons). Who knows what subtle undreamt of possibilities there may be?" (1980, 130). He asks, "Would we say that in such a possible world, it was sound which was light, that these wave motions in the air were light?" He says no, "given our concept of light, we should describe the situation differently" (1980, 130).

I agree. The indicated world does not testify to the genuine possibility of light being pressure waves in the air. But now let us ask a slightly different question. Does it explain the *seeming* possibility of light having turned out to be waves in the air?

[12] I suspect that the explanation is often as simple as this: there is a facsimile of A that might *for all we know a priori* be Q.

Again the answer is no. For that you would need sound to be a facsimile of light. And it is not, for the obvious reason that airwaves do not look the least bit like light. But then what *does* explain the seeming possibility of light turning out to be compression waves in the air? I am not going to comment on that. What we do know is that the explanation is *not* in terms of a genuinely possible strong epistemic counterpart.

One further example, this time not taken from Kripke. Suppose that Q is a broadly geometrical property our concept of which is recognitional. Q might be the property of being jagged, or loopy, or jumbled. It might be the property of "leftiness," which we recognize by asking if the figure in question appears to be facing left (in the manner of 'J' and '3'), or right (in the manner of 'C' and '5'). I will focus for no particular reason on the property of being *oval*. Everyone knows how to recognize ovals, but nobody knows the formula (there is no formula to know). The one and only way to tell whether something is oval is to lay eyes on it and see how it looks. A thing is judged oval iff it looks more or less the shape of an egg.

Now suppose I tell you that *cassinis* are the plane figures, whatever they may be, defined by the equation $(x^2 + y^2)^2 - (x^2 - y^2) = 5$. Is being a cassini a way of being oval? I take it that until you do the experiment, this is an empirically open question. Cassinis could turn out to be oval or they could turn out not to be. You need to draw the figure and see how it strikes you.[13]

This seems not too different, intuitively, from the way LME needs to be sampled to determine whether or not it is heat. Presumably the Kripkean will want to give the same sort of explanation. Just as there are worlds where HME feels hot and worlds where it feels cold, there are worlds where cassinis look egg-shaped and worlds where they look to be shaped like bunny ears or figure-8s.

But this is all a mistake, since for cassinis to look other than egg-shaped to us as we are is impossible. There may perhaps be counterfactual observers who due to their greater visual acuity are bothered by departures from the exact profile of an egg that we ourselves hardly notice. To them, cassinis do not look egg-shaped. But those observers can no more explain the seeming possibility of cassinis' turning out not to be oval than spectrum-inverted observers can explain the seeming possibility of the table's being made of ice. This is because what seems possible (until we do the experiment) is that cassinis look other than egg-shaped to us as we are, with our existing sensory endowment.[14]

****

[13] Cassinis as I have defined them are oval. (They belong to the class of "cassinian ovals"—oddly, most cassinian ovals are not egg-shaped at all.)

[14] It is not as easy as one might think to throw the facsimile requirement over as too onerous. If the appearance that A could be Q is sufficiently explained by noting that *dunce's A* can be Q, then more ought to seem possible than in fact does. It should seem, not only that this brown table could have turned out to be icy, but that it could have turned out to be icy-*looking*, that is, white—for there is (we are assuming) a world where white tables cause the same sort of experience as this brown table causes in me. Similarly the Eiffel Tower should seem like it could have turned out to be three feet in height. For again, a reduced Tower should present to similarly scaled-down observers the same narrow appearances as I enjoy of the real Tower here.

What is the bearing of all this on Kripke's arguments against the mind-body identity theory? Kripke holds that any supposed identities between mental states and physical ones "cannot be interpreted as analogous to that of the scientific identification of the usual sort, as exemplified by the identity of heat and molecular motion" (1980, 150). This is because the model that explains away contrary appearances in the scientific case is powerless against the appearance that pain can come apart from c-fiber firings. Which is more plausible, that the model should suddenly meet its match in illusions about pain and c-fiber firings, or that the model fails to explain away anti-materialist intuitions because those intuitions are correct?

This argument rests on a false assumption, namely that dualist intuitions, if mistaken, would be the sole holdouts against the epistemic counterpart model of illusions of possibility. The model breaks down already in scientific cases like the illusion that this heat could exist without HME (and vice versa).[15] One need not know how exactly the scientific illusion arises to suspect that a similar mechanism might be behind the corresponding illusion about pain.

I do not say the cases are analogous in every respect. The disanalogy stressed by Kripke is this: Identity theorists about heat can concede the existence of a world *v* where HME gives rise to sensations of cold. Materialists cannot, however, concede the existence of a world *w* where c-fiber firings are not felt as pain, because not to be felt as pain is not to *be* pain.

But this puts the materialist at a disadvantage only if we assume that *v* is what it takes to explain why this cold seems like it could have been HME, and *w* is what it takes to explain why this non-pain—this pleasure, say—seems like it could have been c-fiber firings. And my claim has been that intuitions like this cannot be explained by *v* and *w* at all—*unless* their HME and c-fiber firings are such as to feel the relevant ways to us as we are.[16]

The materialist may seem still at a disadvantage, for the following reason. How otherworldly HME feels, we know. It feels hot. But whether otherworldly c-fiber firings are bound to present as pain is not clear. Certainly if they *are* pain, then insofar as it is essential to pain to feel a certain way, that is how c-fiber firings are bound to feel. But what if we suppose with the dualist that c-fiber firings are not *identical* to mental states but *cause* them? The c-fiber firings in *w* might affect minds (ours included) differently than the c-fiber firings here.

I think we should grant Kripke that a world like *w*, *if it existed*, would explain the dualist intuition, at the same time as it verified that intuition. But that is just to say that the intuition would be well explained by *w* if it were correct, which does nothing to show that it is correct. The premise Kripke needs is that we *still* find ourselves with reason to postulate *w* even if we suppose for reductio that it is the identity theory that is correct; this is what supposedly makes materialism a self-undermining position.

[15] One doesn't notice this because Kripke lowers the bar, dropping the facsimile requirement at precisely the point that it threatens to make a counterpart-style explanation unavailable.

[16] Of course there may be other reasons to think *v* exists, e.g., the well-attested phenomenon of the same stimulus causing different perceptual reactions in different perceivers. There are not to my knowledge any well-attested phenomena to suggest the possibility of a world like *w*.

But the stronger premise, we have seen, is false. This suggests to me that Kripke's argument is not in the end successful.

Does this make me a pessimist about conceivability evidence? Not at all. It does put me at odds with

> (O) carefully handled, conceivability evidence can be trusted, for if impossible *E* seems possible, then something else *F* is possible, such that we mistake the possibility of *F* for that of *E*.

But although this was called the optimistic thesis above, a better term might have been super-optimistic or Pollyannaish—because for a type of evidence to *never* mislead about its proper object (the real possibility confusedly glimpsed, in this case) is exceedingly unusual and perhaps unprecedented.[17] The thesis we want, I think, is that

> (O′) carefully handled, conceivability evidence can be trusted, for when impossible *E* seems possible, that will generally be because of distorting factors that we can discover and control for.

Kripke's first great contribution to conceivability studies was to have seen the need for a technology of modal error-detection in the first place. His second great contribution was to have made a start at developing this technology. There is no need to foist on him a third "contribution" of identifying the one and only way modal illusions can arise.

References

Berkeley, George (1979). *Three Dialogues between Hylas and Philonous* (Indianapolis, Indiana: Hackett Publishing Company).

[17] Berkeley suggests a similarly Pollyannaish thesis about perception in *Three Dialogues Between Hylas and Philonous*.

> *Hylas*: What say you to this? Since, according to you, men judge of the reality of things by their senses, how can a man be mistaken in thinking the moon a plain lucid surface, about a foot in diameter; or a square tower, seen at a distance, round; or an oar, with one end in the water.
>
> *Philonous*: He is not mistaken with regard to the ideas he actually perceives; but in the inferences he makes from his present perceptions. Thus in the case of the oar, what he immediately perceives by site is certainly crooked; and so far he is in the right. But if he thence conclude, that upon taking the oar out of the water he shall perceive the same crookedness . . . he is mistaken . . . his mistake lies not in what he perceives immediately and at present, (it being a manifest contradiction to suppose he should err in respect of that) but in the wrong judgment he makes concerning the ideas he apprehends to be connected with those immediately perceived (3rd Dialogue).

Where the Kripkean super-optimist treats seeming failures of imagination as failures of interpretation, the Berkeleyan one shifts the blame rather from experience to inference. The insistence that there are severe, a priori discoverable, limits on our liability to make mistakes about a subject matter often goes hand in hand with idealism about that subject matter. This seems to me a further reason not to associate Kripke with the super-optimistic thesis (O).

Della Rocca, Michael (1996). "Essentialists and Essentialism," *Journal of Philosophy* 93: 186–202.

——(2002). "Essentialism versus Essentialism," in Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Oxford University Press): 223–52.

Hill, Chris (1997). "Imaginability, Conceivability, and the Mind-Body Problem," *Philosophical Studies* 87: 61–85.

Kripke, Saul (1977). "Identity and Necessity," in Stephen P. Schwartz, ed., *Naming, Necessity, and Natural Kinds* (Ithaca: Cornell University Press).

——(1980). *Naming and Necessity* (Cambridge, MA: Harvard University Press).

Nagel, Thomas (1974). "What Is It like to Be a Bat?," *Philosophical Review* 83: 435–50.

——(1986). *The View from Nowhere* (Oxford: Oxford University Press).

Yablo, Stephen (1993). "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research* 53: 1–42.

——(2002). "Coulda, Woulda, Shoulda," in Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Oxford University Press): 441–92.

This page intentionally left blank

Index

1-intension 9, 60–9, 72–3, 78, 104–5, 107–9, 114, 115 n., 124, 130, 136, 201, *see also* intension
2D-intension 144, 160, 163, *see also* two-dimensional intension; intension
2-intension 60–3, 86–7, 115, 120, 127, 130, 135, 137, *see also* intension
'□'-modalization 145, 148, 150, 152, 160–1
Åqvist, L. 311 n.1
accommodation 16–17, 33
acquaintance (direct) 191–5
actuality-based / actuality-neutral hypothesis 333
actuality-based / actuality-neutral modal condition 335
actuality-independence 211–12
actuality in the background of a modal judgement (Yablo) 330
actuality in the content of a modal judgement (Yablo) 329–30
'Actually'-operator ('A') 3, 13, 123, 141 n.1, 142–3, 145, 147, 151, 153, 156–7, 160, 162, 172, 314 n.8
a-intension 62, 129–30, 145, 199, 297, 300–2, 304, 306–8, 313 n.5, 320 n.14, *see also* intension
a-involving expressions 125–7
a priori
  a priori role 14, 38–40, 43
  a priori truth operator 303 n.12
  absolute view of propositional a priority 311, 319–21, 325
  and analyticity 181, 187, 201
  and basic moral principles 220–248
  contentually a priori (Peacocke) 20, 226–30, 233–44
  context-dependent notion of the a priori 197–9, 303 n.12
  and empirical defeasibility 155–56, 159, 163, 187, 195–6
  and epistemic necessity 99–100
  and the epistemic status of self-ascriptions 200–2, 223–8, 233
  and introspection 99–100
  judgementally a priori (Peacocke) 224–29, 233, 234 n.18
  local notion of 9, 197, 306
  propositional a priority 311, 320–1, 325
  relative view of propositional a priority 311, 321–2
  and the skeptical hypothesis 99
  stipulative conception of 100
a-proposition 130, 301, 315, 317, 320–1
anaphora 16–17, 22–36
  and binding 23–4
  discourse anaphora 260 n.3
  dynamic accounts 22, 27–8, 31
  non-dynamic accounts 22–36
  Stalnaker's non-E-type pragmatic account 22, 26, 31–6
  see also E-type accounts of anaphora
assertion 5, 16, 32–3, 259, 263, 265, 269–70, 294
associated proposition strategy 310, 317, 319

Baldwin, Thomas 145, 160 n.24, 168 n., 239 n.26
Bar-Hillel 311 n.1
Barwise, J. 157 n.20
Bealer, George 87 n., 134 n.36, 187
Beaney, M. 251
Bencivenga 195 n.
Berkeley, George 344 n.
Biconditional Fallacy (Peacocke) 248
bivalence 195 n.
Blackburn, S. 232–4, 236–9
Block, N. 66 n., 70 n., 90 n., 137
Bonjour, Laurence 187, 229 n.13
Braddon-Mitchell, David 167
Braun, D. 52 n., 118 n.25
Breheny, R. 16–18, 22 n.1, 26–7, 31
Brown, C. 135 n.41
Burge, Tyler 155, 187, 193, 195 n.
Burgess, John P. 181 n.1
Byrne, A. 14–15, 41 n.2, 92 n., 135 n.40, 198, 199

canonical description 78, 86, 100, *see also* canonical language
canonical language 304
Caplan, Ben 281 n.15
Carnap, R. 55–59, 61, 137, 181, 328
cassinis (example) 342
Categorical Imperative 220 n.1

centered worlds 60–3, 66–8, 71–3, 76, 82–3, 85, 87–9, 93, 103, 105, 108, 110, 112, 115–6, 127–8, 131, 134–7, 208 n., 316
C-fibers (example) 132, 328, 343
Chalmers, D 3, 9–10, 13–16, 40–5, 48–53, 62–3, 65, 78 n., 80, 82–3, 85 n., 89, 90 n., 91–2, 97–8, 103 n.18, 106–7, 109, 115 n., 127–9, 134 n.38, 144–5, 150 n., 185, 198, 201–2, 207, 212, 281 n.16 n.18, 284, 291 n.34, 293, 300 n.5 n.6, 302 n.11, 303–4, 306, 310, 312 n.3, 313 n.5 n.7, 314–20, 322, 324, 331 n., 332 n.
character 2–3, 8, 20, 62–3, 67, 115–9, 136, 145n, 197, 199–200, 281–2, 283 n.23, 284–5, 286 n.28 n.29 , 287, 298–9, 300–1, 324
Cicero / Tully (example) 46–7, 49, 51, 68, 122, 125, 162, 229, 305, 321–2
c-intension 62, 130, 144, 297, 300–2, 304, 306–8, 315, 320 n.14, *see also* intension
cognitive access 302, 307
cognitive significance 39, 43, 55–8, 60–1, 63, 69, 73, 99, 105, 114, 117, 118 n.24, 199
cognitive value (of expressions and thoughts) 199, 300–2, 306
Colterjohn, J. 314 n.7
compatible sentences 84, 92
  epistemically compatible sentences/descriptions 85, 92, 103
  epistemically compatible thoughts 96
complete neutral description, *see* neutral description
Compositionality Constraint (Fodor) 253
Compositionality of reference 20, 254–6
conceivability evidence 327, 344
  optimism about 327–8, 331, 334, 344
  pessimism about 327–8, 244
concepts
  deferential concepts 19, 215, 252–3
  demonstrative concepts 250–2
  epistemic individuation of concepts 253–6
  indexical concepts 19–20, 249–254
  as mental files 250
  natural-kind concepts 252
  past demonstrative concepts 251
  recognitional concepts 19, 251–2, 254
  see also phenomenal concepts; propositional concept
conceptual analysis 3, 293
confirmation 229
Constituency (of concepts) 254, 255 n., 256 n.
constructivism (in ethics) 245
content
  actuality-based modal 331, 333
  actuality-neutral modal 333
  descriptive 11–12, 69, 122, 132
  Evans' notion of 120–3, 153
  Fregean 14, 59, 64, 69, 111, 127, 130, 132
  indirectness of 307
  mental 64, 97, 135–6, 157–9, 166, 300–1, 324
  narrow 64, 70n, 74, 106, 129, 135 n.41, 136–7, 300, 304
  notional 136n
  object-dependent 192, 195, 200
  object-individuated 192
  official 187, 189–91, 194, 196
  propositional 62–3, 112–13, 295, 299, 312
  quasi-Fregean 110, 127, 130
  Russellian account of 261–3, 267
  semantic 113–4, 138, 280, 301, 311, 322
  semantic constraint on 122–3
  singular 190–3, 195, 200
  wide 106
context dependence 9, 64–5, 98, 115, 137, 145 n., 171, 179, 196–7, 201, 303
context-dependent expressions 96–7, 171, 178, 199, 201, 293, 298, 300; *see also* context dependence
context set 5–6, 9, 16, 32–4, 263, 265, 269–70
context shifting operators context dependence 13, 146–8, 171
contextual evaluation 81, 117
contextual intension
  cognitive contextual intension 73, 75, 110–11, 113–14, 128, 130
  cognitive-role contextual intension 73, 111
  conceptual contextual intension 73–4, 137
  epistemic contextual intension 110–11
  evidential contextual intension 75, 131–2
  of an expression type 66
  extended contextual intension 72–5
  fixing contextual intension 75, 134
  functional contextual intension 75, 136
  hybrid contextual intension 71
  intention-based contextual intension 73
  linguistic contextual intension 67–71, 111–2, 115–7, 119, 121, 128, 135, 137
  orthographic contextual intension 67, 70–1, 108, 113, 128, 137
  phenomenal contextual intension 75
  physical contextual intension 75
  physical-phenomenal contextual intension 75, 101
  qualitative contextual intension 133
  semantic contextual intension 68–70, 74, 109–11

token-reflexive contextual intension 71–2, 112–13
see also intension
context-world 312–13
contingent a priori 2–4, 6, 8–9, 120, 124–6, 141, 153–4, 156–7, 159–60, 162–3, 173, 184, 191, 195, 282–3
Continuum Hypothesis 83, 88
context dependence 9, 64–5, 98, 115, 137, 145 n., 171, 179, 196–7, 201, 303
conventional implicature 11, 188
conversational rules 297
Cooper, R. 22, 24
cordate / renate (example) 56–7, 90
Core Thesis (Chalmers) 9–10, 13–14, 64–5, 67–72, 74–5, 77, 81, 83, 89, 104–5, 110, 112, 132, 201, 302 n.11
counterfactual evaluation 100
counterfactual extension, *see* secondary intension
counterpart relation 101, 263, 266–8
Crossley, J. N. 123, 142 n.2, 176,

Daniel / O'Leary (example) 38, 267–9, 294, 296–7, 303
Davies, Martin 3, 8–9, 13–14, 62–3, 121 n., 123–7, 141 n.1, 143–4, 145 n., 146–7, 148 n.11, 149, 152–3, 160 n.24, 168, 171–2, 176 n.3, 178, 179 n.7 n.11, 227, 237, 239 n.25, 282 n.22, 283 n.24, 310
deferential uses 51, 109, *see also* concepts, deferential
Della Rocca, Michael 333–34
demonstratives
complex demonstratives 6 n.3, 11–12, 188–90, 192, 196–7, 199
multi-propositional view of complex demonstratives 188 n.12
simple demonstratives 190, 197, 199 n.23
Dennett, D. C. 135 n.41, 216 n.10
de re attitudes 194–5
de re belief 191–2, 194
Descartes, R. 224
descriptive names 3, 13–14, 121–2, 159–63, 165–73, 176, 179
descriptivism 130, 164–7, 173, 272
argument from error 46–9, 51
argument from examples 45–6
argument from ignorance 46–9, 51
causal descriptivism 301
epistemic argument 164, 166–7
global descriptivism 11, 304–8
modal argument 164–167
semantic argument 164–6
see also holism problem for global descriptivism, indirectness problem for global descriptivism
Dever, Josh 188 n.12
diagonal intension 104, *see also* diagonal proposition; intension
diagonal proposition 4–10, 13, 18, 32–5, 62–3, 67, 112–15, 145, 184–7, 189–90, 196–202, 259, 265–6, 269–70, 297–9, 301, 303 n.12, 304, 313
diagonal reading (Peacocke) 20, 235–8; *see also* diagonal proposition
diagonalization (Stalnaker) 17, 32–6, 297–9, 301 n.9, 304, 308, 311, 312, 314
D-intension 144–5, 160, *see also* intension
direct reference theory 282 n.21, 295, 301, 323–5
Discourse Ethics 248
disquotation, principles of 277–81, 289
D-necessity, *see* necessity
Doggett, Tyler 229 n.1
Donnellan, Keith 169, 191, 195, 321,
double-indexed evaluation 123–4
Dretske 336
dthat 281 n.21, 283
dualism 216, 218, 329; *see also* identity theory
dual-proposition problem / objection 310–11, 316–17, 320–2, 324–5
Dummett, M. 179,
dunce's heat (example) 339, 340, 342

epistemic basis 88, 93
epistemic counterpart model 337–8, 343
epistemic dependence 9, 64, 75–6, 81, 201
epistemic equivalence (or cognitive equivalence) 153–4
epistemic evaluation 81, 89, 104, 121, 129, 134
epistemic equivalence 120–2
epistemic intension 40–5, 48–53, 75, 77–81, 83, 86, 88–9, 91 n., 93–8, 101, 104–12, 114–26, 128–30, 134 n.36, 136–7
applications 105–107
and ascriptions of belief 106
and compositionality 94–5
and contextual intensions 107–112
and deep necessity 120
and deferential uses 108
definition 77–8
and Evans' notion of content 122
and Evans' notion of verification 122
and Fregean sense 106
and Frege's constraint 106
and indicative conditionals 107
and its linguistic expression 48–9, 51
and language free scenarios 107

epistemic intension (*cont.*):
and narrow content 106
and standing meaning 97
of subsentential expressions 93–95
and the diagonal intension 104
and the link between conceivability and possibility 107
and the modes of presentation 106
and the type / token distinction 95–8
and thoughts 96–7
and utterance meaning / content 98
see also intension
epistemic necessitation 78–80
epistemic necessity, *see* necessity, epistemic necessity
epistemic operators 15–16, 285–7
Epistemic Plenitude 84
epistemic possibility 12, 41–4, 48–51, 75–81, 83–84, 88, 92, 95–6, 100, 102–3, 105–6, 119, 129–30, 134, 136–7, 278, 290, 291 n.34, *see also* scenario
epistemic space 75, 93, 103n, 304, 308, *see also* epistemic possibility
epistemically complete sentence / description 84–6, 88–9, 92–4, 100, 105
epistemically invariant expressions 84, 97
epistemically possible claim 76
equivalent sentences 84
essentialism 134, 181–2, 272
E-type accounts of anaphora 17
linguistic 22–7, 30
pragmatic 22–4, 26, 28–31, 36
evaluation circumstances 312–14
Evans, Gareth 3, 9, 13–14, 24–5, 28, 62–3, 77 n., 120–7, 130, 141–2, 145–54, 156, 158, 160–1, 164, 167–73, 176 n.1 n.2 n.3, 177 n.4 n.5, 178 n.5, 179 n.7 n.8 n.10 n.11, 192, 194, 201, 224–5, 228, 233, 250 n., 251, 268 n.16, 307–8
existential generalization 194, 195 n.
extension
A-extension 129, 314–15
C-extension 130, 314–15
one-dimensional 312–13
two-dimensional 312–13

facsimile 337, 339–43
Feynman (example) 46–7, 51, 91
FA-intension 124–6, *see also* intension
'$\mathcal{FA}$'-modalization 145, 148, 152
FA-necessity 124–6
Field, H. 155
Fine, Kit 192 n.
Fixedly Actually (Davies and Humberstone) 13, 227, 237–8, 240–1
'Fixedly'-operator ('$\mathcal{F}$') 3, 123, 143, 146
Flanagan, O. 215 n.
Fodor, J. A. 20, 137, 250, 253–5, 306 n.16
fool's cold 339
fool's gold 337
fool's heat 339–40, *see also* dunce's heat
Forbes, G. 13, 143 n.4, 145 n., 156 n.18, 157–8
free logics 195 n.
Frege, G. 38, 55–6, 58–9, 137, 157, 199, 251
Frege role 14, 39–40, 43, 52–3
Fregean sense 55, 58–60, 68–9, 106, *see also* content, Fregean

García-Carpintero, M. 11, 181 n.1, 188 n.8,
Geach, P. 23, 31
Geirsson, Heimer 321, 324
Generalized Rationalist Thesis 229 n.13
Gertler, Brie 218 n.
Geurts, B. 30
Gilliom, L. 35
Glanzberg, Michael 188 n.12
Gödel (example) 47–53, 91, 97
Goldbach conjecture (example) 278 n.
golden triangle 9, 55, 58, 64, 75, 105, 137–8, 201
Goliath (example) 333–4
good-points 312
Gordon, P. 35
grasping a property 208–12
Gregor, M. 220 n.1
Grosz, B. 35
guarantee
epistemic notion 156–8
modal notion 156
guise of a proposition 324

Habermas, J. 248
haecceitism 103
Hansson, Bengt 311 n.1
Hardin, Clyde 216 n.10 n.11
Hazen, A. 142 n.2
Heat / molecular motion (example) 12–13, 90, 131–3, 134 n.37, 273, 338–43
Heim, Irene R. 24, 260 n.3, 321
Herman, B. 220 n.1
Hesperus / Phosphorus (example) 1–2, 5–6, 8, 42–3, 45, 56, 58–62, 64, 76–7, 79, 86, 90–1, 98, 99 n.15, 101–2, 114, 135, 182–8, 190–1, 197, 199, 273–7, 280, 281 n.15, 294–9, 318–20, 322–4, 327, 331–3
H-intension *see* horizontal intension
H-necessity 149, 153, 173

holism problem for global descriptivism 11, 306–7
horizontal intension 144–5, 160 *see also* intension
Humberstone, Lloyd 3, 8–9, 13–14, 62–3, 123–7, 141–9, 152–3, 160 n.24, 168, 171–3, 176, 179 n.7 n.11, 227, 237, 239 n.25, 282 n.22, 283 n.24, 310
hyperintensionality 210

Identity
  discourse internal identity (Spencer) 259–61, 264, 267 n. , 270
  identity statements 2–3, 14, 18–19, 56–9, 68, 94–5, 97, 125, 162–3, 167, 264–6, 186, 199–200, 273–4, 294–5, 308, 324
identity theory 133–4, 215–218, 328, 343, *see also* materialism; dualism; physicalism
illusion of possibility 1–2, 5, 9, 12–14, 182, 186, 200, 329, 331–2, 334 n.7, 338, 340, 343
  actuality-based illusions 333
  the *if-actually* account of 331–3
  see also Psychoanalytic Standard
implicit conception 21, 226 n. , 241–4
implicit knowledge 165–6, 209–10
indicative conditionals 77, 80, 107, 129 n.
indirectness problem for global descriptivism 11, 307–8
inheritance of epistemic properties 255, 256 n.
Initial Thesis (Peacocke) 221–3, 228
intension *see* 1-intension, 2D-intension, 2-intension, a-intension, c-intension, contextual intension, diagonal intension, D-intension, epistemic intension, FA-intension, horizontal intension, one-dimensional intension, primary intension, Ramsey intension, secondary intension, subjunctive intension, two-dimensional intension, vertical intension
interpretations of the two-dimensional framework
  contextual 9–12, 64–5, 134, 145, 201, 303–4, 313–14
  double-indexing approach 123
  epistemic 9–12, 64–5, 75, 94, 134, 145–6, 201–2
  formal approach 126
  generalized Kaplan paradigm 199, 300–3, 306, 308
  metasemantic 114–15, 145–6, 184, 186–7, 196–7, 301–4, 308
  neo-Fregean 11–12, 190–200
  semantic 114–15, 123, 130, 196–8
  simple modal conception of 145, 159

Jack the ripper (example) 98, 111–2, 118 n.24,119
Jackson, Frank 3, 9, 14, 19, 41 n.2, 42, 44, 49, 51 n., 53, 62–3, 78 n., 89, 91–2, 129–30, 144–5, 163, 165–6, 206, 281 n.16 n.18, 293–4, 297, 300 n.5 n.6, 304, 306, 310, 313–17, 319–20, 322, 324,
Jeshion, Robin 191–3, 195–6
Johnston, M. 232 n.16, 247
judgementally valid 224–8
Judson, Whitcomb L. (example) 160–2, 168, 171–2, *see also* Julius
Julius (example) 3, 113, 120–2, 124, 160–2, 167–72, 176, 178–9, 307–8

Kamp, Hans 28, 260 n.3, 311 n.1,
Kant, I. 55, 58, 220, 222, 240–1
Kaplan, D. 2–3, 8–9, 13, 62–3, 67, 115–19, 123, 127, 145n, 188, 197, 199, 200–1, 227, 249, 281–2, 293, 298, 300–4, 310, 311 n.1, 313 n.7, 314 n.7 n.8, 323, 342
King, Jeffrey 188 n.10
Korsgaard, C. 232
Kripke, Saul 1–6, 9–12, 14–15. 40–1, 46–52, 55–6, 58–61, 100–102, 114, 120, 127, 130–5, 157 n.19, 158 n., 164–5, 173, 181–3, 187 n., 190, 196, 201, 227, 272–81, 284, 289–90, 294, 296, 310–11, 315–16, 318–21, 327–29, 331, 333, 334 n.6, 335–44
Kripke's noncircularity requirement 52
Kroon, F. 165

Lehmann, Scott 195 n.18
Leibniz, G. 220
Lenhardt, C. 248 n.38
Leverrier (example) 95, 97, 109, 192, 306
Lewis, C. I. 57 n.2, 230 n.14
Lewis, David 11, 57 n.1, 127, 135–6, 146, 182–3, 198, 232 n.16, 239 n.25, 262 n.8, 272 n.4, 273, 281 n.17 n.18, 284, 304–6, 310, 311 n.1, 317
Loar, B. 136 n.41, 214 n.
Locke, J. 220
L-truth 57 n.2
Lumpl (example) 333–4

Macià, Josep 198 n.
MacIntosh, D. 314 n.7
Marianna (example) 206, 216
Mars (example) 4, 15, 60, 101, 183–4, 300, 332
Mary (Frank Jackson's example) 19, 206, 215–18
materialism 83, 328, *see also* dualism; identity theory; physicalism
McGlone 280

McKinsey-style reasoning 192–3
McLaughlin, B. P. 192
meaning-constitution point 187, 189–90, 196
Metaphysical Plenitude 82–3, 85, 88–9, 103, 105, 107, 129
  counterexamples to 82–83
metaphysical possibilities 76, 81, 83, 103, 105, 278, 281, 283, 290, 291 n.34
Michael, Michaelis 319 n.12, 321–3
Millianism 185, 197–200
mind-body identity theory, *see* identity theory
mind-body problem 201, *see also* identity theory
mind-dependence 220–1, 230 n.13, 232–40, 242, 244–8
modal equivalence 164–6
modal illusion, *see* illusion of possibility
modal rationalism 103 n.18
mode of presentation 64, 98–99, 106, 159, 295, 324, *see also* Fregean sense
moderate rationalism 21, 221, 238, 241–2
Montague, Richard 260 n.4, 311 n.1
Moore, G.E. 239–40
moral principles 20–1, 220–3, 228, 230–41, 244, *see also* Initial Thesis (Peacocke), Sharpened Thesis (Peacocke)
moral rationalism 21, 220–48

Nagel, Thomas 328–9
natural properties 305
Neale, S. 24–5, 28
necessarily coextensional concepts / properties 214, 216–18
necessary a posteriori 2–3, 5–6, 9, 11, 45, 114, 120, 125, 135, 141, 153, 157, 161–4, 167, 184, 196, 272–91, 310–11, 315–19, 321, 323, 325
necessity
  de dicto 182
  deep necessity / contingency 120, 123, 149–54, 154, 156–7
  de re 3, 9, 181–3, 200
  D-necessity 149, 152–4, 159–60, 163, 173
  epistemic necessity 78–81, 96, 98, 100, *see also* a priori
  H-necessity 149, 153, 173
  metaphysical necessity 227, 229, 240
  propositional necessity 320
  quasi-necessity 313, 316
  superficial necessity / contingency 149–50, 152–4, 156–63, 165, 173
  superficial versus deep contingency and necessity (Evans) 3, 13–14, 141, 149–67, 172–3, 178
Nelson, M. 51 n.11
Neo-Fregean Thesis 58–9, 61, 64, 106
Neptune (example) 42, 95, 97, 109, 192, 306
neutral description 86–9, 103, *see also* canonical description
neutral language 304, 308, *see also* canonical language
Nicholsen, S. 248 n.38
Nida-Rümelin, Martine 16, 18–19, 206 n.2, 207 n.4, 214 n.7, 216 n.10 n.12, 219 n.
non-introspective way of coming to judge 225–6, 228, 233
notional worlds 135 n.41

one-dimensional intension 144–5, 304, 313, *see also* intension
Oscar / Twin Oscar (example) 67–9, 74, 109
ostensive definitions 47, 279 n.13

pain (example) 132–3, 134 n.37, 223–224, 230, 236–7, 243, 328–9, 343
Partee, B. 23
Peacocke, Christopher 3, 20–1, 155 n.17, 187, 221 n. , 226 n. , 229 n.12 n.13
Pérez Otero, Manuel 181 n.1
Perry, John 32, 35, 157 n.20, 188, 250, 262 n.7, 324
Phenomenal character 212–13
phenomenal concepts 18–19, 205–19, 328–9
  actuality-independence of phenomenal concepts 212–13
  functionalism about phenomenal concepts 209
Phenomenal properties 18–19, 205–19
phenomenology 41, 111, 336–7
Phosphorus (example) 1–2, 4, 7, 15, 42–3, 56, 58–61, 64, 76–7, 79, 90, 98, 99 n.15, 114, 135, 182–5, 188, 197–8, 273–7, 280, 281 n.15, 294–9, 318–20, 322–4, 327, 331–3
physical-functional concept and phenomenal concepts 215–18
physicalism 40, 88, 92, *see also* identity theory
Pierre (example) 279, 299
Plantinga, Alvin 319
Plenitude Principle 81, 84–5
Pollyannaish 344
possession-conditions for concepts 242–3, 245–6, 253–5
possible centered world, *see* centered world
possible worlds, *see* centered world; epistemical possibilities; metaphysical possibilities
presupposition 5, 6 n.3, 16, 18, 26–7, 32–35, 188, 258–67, 270
Prichard, H. 243
primary intension 15, 62–3, 127–9, 145, 282–7, 289, 315, 318

primary intension of a concept 208, 210, 212
see also intension
Principles of Possibility (Peacocke) 241–4
proposition
associated 115, 135, 310, 318–19
diagonal reading (Peacocke) 20, 235–8
enriched 280
expressed 62, 115, 120–1, 135, 272, 275, 278–80, 281n., 282, 283 n.23, 284–6, 295–8, 302–3, 307, 310, 313, 315, 317–19, 321–3
indexical proposition (Perry) 32
as sets of possible worlds 112, 135, 272, 281, 283, 286, 289, 316–7, 318 n., 324
singular 62, 98, 115, 187, 189, 191, 194, 275, 283 n.23, 297, 300, 302, 307–8, 323–5
structured 62 n, 65
see also diagonal proposition, horizontal proposition, one-dimensional proposition, two-dimensional proposition
propositional concept 3–8, 11, 19, 112, 184–8, 190, 196–7, 200, 295–9, 301, 303–4, 312–13, 324
proprioception 250
Pryor, J. 14–15, 41 n.3, 92 n., 155, 196 n.20, 198–9
Psychoanalytic Standard 12, 334–6
psychoanalytically acceptable explanation 335, 339–40, see also Psychoanalytic Standard
Putnam, H. 15, 40, 45–46, 101, 127, 252, 305
Putnam's paradox 305

qualitatively identical situations 87 n., 103, 131–3, 134 n.36, 274, 276, 278
qualitatively identical statements 131–4
quantificational modal logic 181
Quine, W. V. 181, 260, 272, 312 n.3, 316 n.

Ramsey 305
Ramsey intension 80–1, *see also* intension
Ramsey sentence 92
Ramsey test 77, 80
rational significance, *see* cognitive significance
rationalism 221 *see also* modal rationalism, moderate rationalism, moral rationalism
Rawls, J. 220 n.1
realism about possible worlds 182–3
Recanati, F. 16, 19–20, 250 n., 252
reference-fixing authority 39–40, 47, 53
reference-fixing descriptions 7, 160, 165, 184, 311, 316, 337–8
reference-fixing descriptive presupposition 197
reference-fixing information 5, 7, 14, 198–9
reference-fixing role 14, 39–40, 43, 47–9, 52–3
Reichenbach, Hans 119, 187, 193, 201–2
response-dependent concepts (Johnston) 247
Richard, Mark 324
rigidified description, *see* rigidity
rigidity 147, 179–80, 323–4
deeply rigid designator 163
rigid designators 1–2, 59, 61, 86–7, 102, 109, 132–3, 134 n.36, 181–2, 187–90, 272–3, 291 n.33, 296, 299–300, 307, 311, 313, 316, 323–4
rigidification 126, 306–7
rigidified descriptions 92 n., 130, 283, 285–8, 305–6
superficially rigid designator 163
Russellian singular term 168–73
Ryle, U. 28

Salmon, Nathan 98, 99 n.15, 262 n.7, 279 n.13, 314 n.7, 319 n.12, 321, 324
Sasha (example) 279 n.13
Scenario 30, 41–2, 48, 76–87, 89, 92–5, 97, 100–5, 107–10, 114, 119, 128–9, 134, 136, 209, 274, 304, 334, 336–7
as centered worlds 82
epistemical approach to 85, 105
as epistemically complete hypothesis 84
linguistic construction of 83–5
and maximal hypotheses 81, 83–86
metaphysical approach to 85
as a modal primitive 83
the world-based view 85, 108, 110
Schiffer, S. 106–7
Schmidt (example) 47–50, 52–3
Schroeter, L. 85 n., 103 n.18
secondary intension 15, 19, 62, 127, 144, 282–6, 288–9, 315–16, 318
secondary intension of a concept 208, 210, 213
see also intension; vertical reading
semantic pluralism 65, 105
semantic stability 87 n., 134
semantic theories: descriptive versus foundational 8, 10, 196–7
semantic vs. pragmatic information 259–60
semantically neutral expressions, *see* neutral description; neutral language
semantically stable expressions 87 n., 134 n.36
Sharpened Thesis (Peacocke) 20–1, 228–30, 239–40
Siegel, Susanna 188 n.12
Simpson, Paul 314 n.7
Smith, M. 232 n.16
Soames, Scott 15–16, 50 n., 52 n., 92 n., 98, 99 n.15, 106 n., 135 n.40, 164 n.28,

Soames, Scott (*cont.*):
182–3, 186 n., 190 n., 262 n.7, 272 n.1, 276 n.10, 279 n.13, 280, 282 n.21, 284 n.26, 285 n.27, 289 n.30 n.31, 291 n.34, 324
space of possible worlds 56, 102–3, 123, 304, 308, *see also* epistemic possibilities; metaphysical possibilities; possibility
Speaks, Jeff 291 n.34
spectrum inverted 209, 335–6, 342
Stalnaker, Robert 3–12, 16–19, 22, 26, 30–4, 36, 41, 62–3, 66 n., 67, 70 n., 72 n., 90 n., 112–15, 123 n., 127, 144–6, 182–90, 196–200, 234, 259, 261, 262 n.8, 263, 265–7, 269–70, 272 n.3, 281 n.18 n.19, 284 n.25, 291 n.34, 293n, 294 n.1, 298 n.3, 301 n.7 n.9, 304 n.14, 305 n.15, 310–12, 313 n.5, 317, 324, 332 n.
standing meaning 2, 69, 97–8, 125
Steel Earth / Steel Oscar (example) 67–9, 74, 108–9,
Stoljar, Daniel 329 n.1
Strawson, Galen 215 n.
Strawson, P. 250
Strong Disquotation principle 277, 279–80, 281 n.15, 289
Strong Disquotation and Justification principle 277–80, 281 n.15
strong epistemic counterpart, *see* facsimile
subjective character of sensations 212, *see also* qualitative character; Phenomenal character
Subjectivist Fallacy 221, 244–8
subjunctive intension 100–2, 104, 107, 112, 120, 127, 130, *see also* intension
subjunctive necessitation 101
subjunctively complete sentence 100
supervaluationist semantics 195 n.
sympathetic imagination 328–9

table made of ice (Kripke's example) 131–2, 273–5, 277–8, 291 n.33, 327, 331, 335–6, 338–42
Taylor, B. 157 n.20
Taylor, Richard 330
Thau, M. 53 n.13
Tichy, P. 135, 311, 317–20
thoughts
object-dependent 168–9, 192–3
object-dependent 192–3
thought content 157–9, 166
transparency principle 214–16
truth
minimalism about truth 232
counterfactual theory of truth 195 n.
Twin Earth (example) 42, 67–8, 76, 109, 128–9, 136, 252, 302, 311, 337
Twin-Earthable expressions 86–7, 302, 304
two-dimensional;
function 207, 210–11
matrix 149, 312, 318 n., 324
modal logic 123 n., 141–2, 144, 149, 153, 272, 311, 314
modal operator 303 n.12
modal semantics 293, 303n.
see also interpretations of the two-dimensional framework
two-dimensional intension 61–3, 102, 104, 298, 300–4, 307, 313, *see also* intension
two-dimensional proposition 19, 32, 112, 259, 265, 317, 318 n., 321, *see also* two-dimensional intension
two-dimensionalism
ambitious two-dimensionalism 281 n.21, 282–5, 289
benign two-dimensionalism 281–2
strong two-dimensionalism 106 n., 281, 283–9
weak two-dimensionalism 283–5, 289
Tye, M. 192

underlying possibility 331, 340
Uranus (example) 95, 97, 191–2, 306
Urmson, J.O. 243 n.
utterance problems 9, 13, 70–1, 74, 128, 200–2, 313 n.7, *see also* a priori and the epistemic status of self-ascriptions

van Rooy, R. 30
Venus (example) 7, 42, 56, 59–62, 90, 98, 101–2, 135, 184–5, 191, 193, 275, 294–5, 324, 332
vertical intension 144–5, 149, 173, *see also* intension
vertical reading (Peacocke) 235–7
V-intension; *see* vertical intension
Vision, Gerald 314 n.7
V-truth 91–2
Vulcan 192, 196

Water / H2O (example) 19, 41–3, 45–6, 56, 58–9, 61, 67–9, 74, 76–7, 80, 86, 89–90, 93, 101, 113–14, 128, 130, 252, 307, 310–13, 315–18, 322, 337
watery stuff (example) 67, 311–13, 315–16

Weatherson, B. 107
White, Stephen 136–7, 214 n.
Wiggins, D. 232 n.16, 239 n.24 n.25
Williams, B. 246 n.
Williamson, Timothy 198
Wittgenstein, L. 198, 246
Wong, Kai-Yee 14, 310, 311 n.2, 313 n.6, 314, 319, 321, 324
Wright, C. 232 n.16, 239 n.24, 247

XYZ (example) 41–6, 67, 76–7, 79–80, 82, 86, 101, 109, 128, 136, 252, 311, 313, 315, 335 n., 336–7

Yablo, S 12–13, 77 n., 80, 109

zombies (example) 83, 331 n.