

Chapter 1

Moral Realism, Meta-Ethical Pyrrhonism and Naturalism

Drew Khlentzos

Abstract This paper argues that naturalistic moral realism is vulnerable to a *Hard Problem* that has gone largely unrecognized. This problem is to explain how natural moral properties are detected by the folk. I argue that Thomas Nagel's persuasive case for moral realism founded on the priority of first-order moral evaluations over second-order reflection is not conclusive—a certain type of moral agnosticism which I call Meta-Ethical Pyrrhonism can account for our inability to think of first-order moral evaluations as merely subjective or relative. Although unsatisfactory as metaphysics, Meta-Ethical Pyrrhonism is arguably all that a moral naturalist is entitled to by way of a meta-ethical theory.

Keywords Hard Problem · Meta-Ethical Pyrrhonism · Moral realism · Moral skepticism · Naturalism

Suppose that there are no non-natural facts—that all that exists is determined by ultimate physical laws and boundary conditions. Can any such naturalistic scheme countenance moral norms and values? Mackie (1977) notoriously thought not. Moral norms and values are simply too queer to be part of the furniture of the world, he contended. The reason was that these norms and values as we construe them are supposed to be objectively prescriptive. Yet objective prescriptivity is a 'queer' property so unlike any other natural property that we could never have any naturalistic ground for believing in it.

Mackie's Argument from Queerness has not convinced naturalistic moral realists. Some hold that Mackie is mistaken in thinking that objective prescriptivity, at least as Mackie himself construed it, is a *sine qua non* of folk moral norms and values, others that the very notion of objective prescriptivity rests upon a confusion of ontological with motivational questions.

D. Khlentzos

Language and Cognition Research Centre, School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale, New South Wales, Australia 2351 Tel.: +61-2-67733208 e-mail: dkhlezt@une.edu.au

My primary purpose in this paper is not to offer a substantial contribution to this debate, although I shall need to say something about objective prescriptivity and its place in folk morality and in this context shall explain why Mackie's notion cannot be lightly dismissed by naturalists. My aim is to first examine from a naturalistic point of view the phenomenon with which Mackie starts: the conflict between our ordinary moral judgments and naturalistically grounded reflections on the nature of those judgments.

I shall argue that Mackie is right to find it extremely difficult to bring first-order ethical judgments into harmony with second-order reflection on those judgments. Furthermore, moral naturalists have not appreciated the complexity of the task they face—they have, I shall argue, largely ignored a foundational problem of naturalistic moral epistemology. This will involve a brief discussion of what naturalistic moral realism is and of why it is vulnerable to a difficulty I label the *Hard Problem*. I shall then argue that Thomas Nagel's persuasive case for moral realism founded on the priority of first-order moral evaluations over second-order reflection is not conclusive—a certain type of moral agnosticism which I call Meta-Ethical Pyrrhonism can account for the phenomenon Nagel emphasizes: namely, our inability to think of first-order moral evaluations as merely subjective or relative. Although unsatisfactory as metaphysics, Meta-Ethical Pyrrhonism is arguably all that a moral naturalist is entitled to by way of a meta-ethical theory.

1.1 Objective Prescriptivity and the Requirements of Naturalism

Suppose you hear a voice instructing you to strangle your cat. Checking to see where the voice is coming from you realize, to your dismay, that it has issued from within your own head. Naturally, you do not think to obey. For as you now realize, you have just experienced a transitory psychotic episode. Schizophrenics hear such voices and, tragically, often obey the instruction conveyed, however bizarre it may be. Why do they do so? Given their condition, sometimes because of the tone or the urgency of the instruction, other times because of the authority they believe to have issued it. The schizophrenic who strangles his cat may well do so because he believes that this *had* to be done. If he believes that God or Satan instructed him to do so, he has identified the source of this *to-be-doneness*. On the other hand, if we can get him to acknowledge that he was merely reacting to the urgency of the instruction or the insistent tone in which it was delivered, we may just succeed in convincing him that he was mistaken in believing the action of strangling his cat really was objectively prescriptive.

Any naturalist who believes that it is straightforwardly and objectively wrong to incarcerate young children who have committed no crime owes us a naturalistic account of the source and nature of this objective wrongness—one founded upon natural, ultimately physical, facts. For a number of reasons, moral naturalists seem not to have fully appreciated the complexity of this task. This is partly because of two misplaced prior commitments—to the metaphysical thesis of moral supervenience

and to the belief that central folk moral judgments must be true. The combination of these two has led naturalists to misconstrue their fundamental metaphysical task so that it becomes for them a question of either showing how the totality of natural facts fix the moral facts or, if non-reductive supervenience be judged inadequate, identifying those natural properties to which the moral properties are to be reduced.

However, the fundamental metaphysical task for the naturalist is *not* to vindicate any sort of folk theory, however venerable that theory may be. The fundamental task is to show how, in the moral case, normative moral properties are so much as *possible* in a purely physical world: how does a world of fields and forces give rise to anything so strange as a moral obligation or a virtue or a duty?

To be sure, this is no isolated problem. The moral naturalist's task is part of a wider naturalistic metaphysical task that I've dubbed the *Representation Problem* in earlier work (Khleutzos 2004). The *Representation Problem* is the problem of explaining how mental representation of any feature of a mind-independent world can be possible, ethical and normative features included. Naturalistic realists have been slow to appreciate the centrality of the *Representation Problem* as well.¹

One very plausible naturalistic answer to our question is that a physical world *cannot* give rise to any such things as duties or virtues or values except by giving rise to intelligent creatures who invent them without realizing this is what they've done. This was, of course, Mackie's answer. If there are no objective moral facts or properties to begin with, then there are none to be fixed by the physical facts and there are none that are reducible to those facts. Whence, the naturalist cannot start with the assumptions most moral naturalists start with.

The naturalistic task with respect to folk moral theory must therefore be one of radical interpretation. Moral language is a jungle language, to be interpreted from scratch using *bona fide* natural properties, and not one that the theorist already knows how to speak. Along the way, the naturalist will need to pose and answer questions that might look bizarre to the folk or that might offend common sense, questions such as this: how is the folk appeal to the voice of moral conscience to be distinguished from the schizophrenic's appeal to voices inside his head?

So, consider our example once more and assume the naturalist concurs with the folk belief that incarceration of innocent children is wrong. To the extent our naturalist is a realist, s/he will contend that such incarceration would still be wrong even if we all agreed it wasn't and that our merely judging it to be wrong does not make it so. There is a judgment-independent fact to the matter as to whether incarceration of the young is wrong. Moreover, the confident and unanimous folk belief that it is wrong could at least in principle be mistaken, according to the naturalistic moral realist.

What is there to say it is not? How do the folk manage to detect and respond to the natural property of objective impermissibility inhering within acts of incarcerating the young? The question does not merely ask: how do the folk form beliefs about what is objectively impermissible? It asks how they form *veridical* beliefs such as the belief that incarceration of innocent children is objectively impermissible, assuming it is. And the question is *not*: 'how does the *naturalistic theorist* detect this natural property?' It is: 'how do the *folk* detect this natural property?'

Suppose the natural property of objective impermissibility is its tendency to minimize overall well-being if adopted as a rule for action. Positing a brute faculty of moral intuition in the folk for detecting this property is clearly unacceptable to any naturalist. The suggestion must be that the folk infer its existence somehow.

Infer it from what? Their moral feelings, perhaps? The folk feel disgusted or outraged by the prospect of seeing innocent children behind bars and infer that the best explanation of their disgust and outrage is that these feelings are elicited by the gross-well-being-inhibiting feature of incarceration. But this inference looks hazardous to the extent that it may very well tell against the moral realist. For if child incarceration is objectively impermissible, then the direction of explanation, for a moral realist at least, needs ultimately to proceed in the reverse direction: we feel disgusted *because* it is wrong, rather than, as non-cognitivists aver, we infer that it is wrong simply because we all feel disgusted by it. Moreover, and not insignificantly, this is precisely the explanation the folk themselves give of their attitudes. Rightly or wrongly, the folk believe that they can detect moral properties directly.

Furthermore, for this inferential story to have any hope of working, we'd need to be able to distinguish *objectively* grounded moral disgust or outrage from non-objectively grounded moral disgust or outrage, since there are many things that disgust/outrage some that do not disgust/outrage others and there are even things that tend to disgust or outrage (almost) everyone—such as the prospect of intense suffering of a loved one in the terminal stages of cancer—that are not objectively impermissible. Yet the task of distinguishing the relevant objectively grounded moral feelings from the non-objectively grounded moral feelings simply raises anew the very problem this type of account was invoked to solve: how do *the folk* detect the natural property of objective impermissibility?

Do things improve for the naturalist if, instead of moral sentiments, we appeal to moral judgments as evidence for the existence of a natural property of objective impermissibility? Perhaps, as suggested in Nagel (1997), it is the folk's inability to think of incarceration of children as anything other than objectively impermissible that provides the strongest evidence that their moral judgments track objective properties.

We are going to examine Nagel's view in more detail later but we can note this now: even if Nagel's suggestion is right, it still doesn't solve the moral naturalist's dilemma. Positing appropriate natural properties to identify with folk moral properties is the easy part; saying how untutored folk detect and respond to them in arriving at their moral judgments is the *Hard Problem* of naturalized moral epistemology.

Surprisingly, naturalists seem to have paid little if any attention to the *Hard Problem*. Many write as if there is no obligation on a naturalist to address it. Thus, according to Susan Hurley, 'to say that a certain act *ought* to be done is to say that it is favored by the theory, whichever it may be, which gives the best account of the relationships among the specific values that apply to the alternatives in question' (see p. 11 in Hurley 1989).

Yet the folk who make true moral judgments, such as that incarceration of children is wrong, know nothing of any moral *theory*. If the folk moral judgment and the theorist's moral judgment are not to simply diverge in meaning when each declares

child incarceration wrong, we need some way of connecting folk evaluations to the theorist's. Moreover, the connection between folk 'wrong' and the theorist's 'wrong' had better be a very direct one, otherwise, to quote Bas van Fraassen, the theorist runs the risk of 'giving in to one of the temptations that make for really bad philosophy, surreptitiously and underhandedly turning the ordinary word . . . into technical jargon while ostensibly keeping it intact' (see p. 7 in van Fraassen 2002).²

To underscore the interpretative task facing the naturalist, imagine that our theorist, Nat, has stumbled upon a lost race of people completely cut off from civilization, somewhere in the jungles of Malawi. Remarkably, these people (the Churls as they call themselves) appear to speak English or something that sounds awfully like it. However, their moral vocabulary, if that is what it is, seems to differ from our own. Instead of describing acts as 'right' and 'wrong' the Churls classify them as 'cornucopian' and 'pandoran'. Nat is excited to learn that acts described as pandoran in Churlish can be mapped onto those acts that he knows to have high overall disutility for the Churls and that acts that are cornucopian can be mapped onto acts that possess high overall utility for the Churls. Our theorist is thus satisfied with his discovery that 'pandoran' just means *possessing high overall disutility* and that 'cornucopian' means *possessing high overall utility*.

What if the Churls reject this identification though? What if Nat has failed to appreciate the religious significance of the terms 'pandoran' and 'cornucopian'? What if 'pandoran' in Churlish means *cursed by the gods* and 'cornucopian' in Churlish means *blessed by the gods*? Should Nat *still* maintain that in spite of what the Churlish folk themselves say, he has located the natural properties that are the *real* Churlish moral properties on the grounds that a mature Churlish folk morality freed of its current entanglement with primitive Churlish religion would independently endorse the property identities he has posited?

Of course, our theorist *could* maintain this and could even do so with some plausibility if he could show how the Churls are able to first detect and then respond to overall utility/disutility without realizing (even denying) that this is what they are doing. Yet even this would not suffice. Detection followed by response is necessary but it is far from sufficient. The good folk of Salem were able to detect and respond to the effects of ergot poisoning which they mistook for demon possession, but no one would think of rehabilitating witch theory by suggesting that a mature folk witch theory freed of any entanglement with the primitive fears and superstitions associated with witches would identify the property of demon possession with that of ergot poisoning.³

The allegory of Nat and the Churls demonstrates what is needed in the moral case. The folk must unwittingly detect and respond to the relevant natural moral properties—what *the folk* call 'morally impermissible' is really what *the theorist* identifies as gross-well-being-inhibiting features, say, even though the folk have no inkling that this is what they are detecting and responding to and may roundly reject the suggestion that it is. Let us assume that overall utility is the relevant natural property. The idea then is that moral judgments are no different from water judgments in this respect. Even if we all believed that water was a living spirit, our water judgments would still in reality be H₂O judgments.⁴

But how do the folk manage this for their moral judgments? How do they lock onto the right natural property? Perhaps there is some module in the brain that performs the relevant bulk utility calculations? We would need more than this though. We'd need a *reliable* utility calculator, at least for the relatively uncontentious cases such as the incarceration one. The output of the module must map onto the facts, on pain of failing to vindicate moral realism.⁵

So perhaps reliable utility modules have been selected for in our evolutionary history? There are a number of difficulties with this proposal, not the least being that the empirical evidence attests to a widespread divergence in basic moral evaluations that seem to be closely linked to the moral evaluator's culture. This affords some *prima facie* evidence against the existence of a shared module.⁶ The thing to note at this stage though is that the moral naturalist needs some such suggestion to be true.

Why? In the case of water and H₂O, there was no need to posit a dedicated H₂O-detecting brain module. So why is it needed here? The reason brings us back to Mackie. Moral judgments, as Mackie argued, directly engage the will. Unless the impermissibility-grounding property were to be somehow represented as such in the brain, there would be no naturalistic justification for thinking that it was *that* property rather than some other one that moved the folk to evaluation and action. Yet if our theorist were to discover that the property s/he had identified with the impermissibility of incarcerating children was *not* causally responsible in the folk for their attitudes of disapproval, expressions of disgust and outrage, political agitation and so on, s/he would have to revise the conjectured identity. For *qua* radical interpreter, the naturalist wishes to know what natural property it is that the folk recognize and respond to when they issue their moral judgments and evaluations.

The moral case contrasts sharply with water and H₂O. The folk have no dedicated mental module for detecting the presence or absence of the chemical property *being composed of H₂O*. Nor do they need any. It suffices, for the purposes that water serves for individual survival and social transactions, that there be some semantic marker for the substance water acquired in the ordinary process of learning the significance of that substance. It is unnecessary that there also be some pre-coded mental formula representing the composition of that substance.

For the moral case though, no physical substance is the subject of moral predication whence the substance/composition distinction lapses. Thus, if the property of being objectively impermissible is to be identified with the natural (dispositional) property of leading to a much lower ratio of pleasure to pain, and if, unbeknownst to the folk, it is *that* property which they actually detect when they evaluate child incarceration as impermissible and which, once detected, leads them to decry as immoral the actions of any political regime instantiating it, then they had better have some reliable means of detection of that property. Moreover, this means of detection would need to be relatively encapsulated and not subject to top-down processing, if it is to be compatible with widespread ignorance as to its true nature. That suggests a dedicated brain module.

More complex hypotheses are possible of course. There is no reason why a moral naturalist should not be a pluralist about moral values and attribute a variety of different such modules suitable for detecting other natural properties

that could serve within different individuals as value properties. Alternatively, the modularity hypothesis might be challenged by some moral connectionist model. These are details that need not concern us.

Mackie of course did not posit anything like a module on behalf of moral naturalists, since he wrote many years in advance of the modularity of mind thesis. What he (famously) argued was that objective values would need to somehow attract the will. Yet some naturalists who are externalist in their theory of moral motivation reject this requirement as gratuitous, whilst others believe that it is possible to embrace less extreme versions than Mackie's on the question of objective prescriptivity and the will. In fact, as we shall see, Mackie's requirement seems entirely justified once one appreciates the Hard Problem. But let's review the objections to it first.

In a review of Susan Hurley's *Natural Reasons*, Frank Jackson writes:

If Mackie is right, if it is an essential part of **M** (matured Folk Morality) that there be properties which of necessity attract anyone who contemplates them, no doubt nothing is (morally) right. But more moderate views about **M** are possible (see p. 486 in Jackson 1992).

To anyone who appreciates the Hard Problem of saying how it is that the folk first detect and then respond to those natural properties that are the moral properties, the requirement that the relevant natural properties of necessity 'attract the will' does not look at all gratuitous. For this is precisely the explanation *the folk* themselves tender when quizzed about why they acted in the way they did on certain occasions:

Nat "What moved you to report those young guys to the police?"

Subject "They set fire to a cat!"

Nat "But what was it about that action that disturbed you?"

Subject "It was blatant cruelty to an animal! What would you think?"

If our theorist has identified high overall disutility with an action's wrongness, then it must have been a recognition of the high overall disutility of the action of setting fire to the cat that prompted the subject's moral disgust and subsequent action. At least it needs to have been that if the subject's moral evaluation is to be objectively true. This is just to say that it is *that* natural property of that type of action that repels to some degree or other any rational, fully informed moral agent who contemplates it. What 'more moderate' view could Jackson have had in mind? Only if one adopts the highly immoderate view that recognition of moral value has *no effect per se* on the will could one think otherwise.

That leads into the next criticism since the above is in effect the externalist view. Far from it being built into the nature of natural moral properties that they provide the source of moral motivation, it is simply a result of socialization that we feel motivated to care about others to any degree or to act on our moral judgments and evaluations, according to externalists about motivation.

Now there is *something* to this externalist objection, but it is opaque whether it could do anything other than exacerbate the Hard Problem if it were true. That there is a difference between being more motivated to perform action A rather than action B and ranking A ahead of B in some preference order seems a *sine qua non* of any credible account of akrasia. It would be a very odd theory of action though

that explained normal psychological motivation as a mere artefact of socialization, taking as its model for action the pathological case of the akratic. Moreover, the empirical claim that it is purely by dint of socialization that we care about others or feel motivated to act on our judgments seems very dubious at best. There is a mounting body of psychological and neurophysiological evidence attesting to the crucial and ineradicable role emotions play in decision-making and an equally impressive body attesting to the genetic provenance of feelings of empathy towards others and dispositions toward reciprocity.⁷

The crucial point for our purposes though is this: suppose that externalists are right and that moral motivation is just a socially-induced phenomenon. Suppose as before that the morally impermissible is the natural property *tending to reduce the ratio of overall pleasure to overall pain*. Call this natural property R. The Hard Problem of naturalistic moral epistemology is to explain how R determines the content of (or otherwise enters into) folk moral judgments.

Externalists reject Mackie's suggestion that R must somehow engage the will. So we assume that it is the socially-induced internal property S that does this instead. For simplicity, assume S is just the socially sanctioned rule 'Reject incarceration of the young and innocent!' Now when the folk pass the moral judgment that incarceration of children is impermissible, it is no longer R but S that moves them to decry such incarceration and agitate for its removal from political regimes that institutionalize it. For the radical interpreter, what is the point then of appealing to the external natural property R in any naturalistic explanation of why the folk issue the particular moral judgments they issue when, *ex hypothesi*, it is S not R which motivates them to do and say what they do and say? *Externalism about moral motivation simply exacerbates the Hard Problem.*

I think this was one of the reasons why Mackie drew a contrast between objective and subjective values. There were, he argued (Mackie 1977), two independent constraints that moral values needed to satisfy:

1. *They had to systematically and reliably guide decision and action.*
2. *They had to be grounded, independently of the agent, in the objective natural order.*

According to Mackie, whilst objective values satisfied (2), they failed to satisfy (1). Subjective values, on the other hand, satisfied (1) but failed to satisfy (2). In order to ensure that objective values, such as the moral naturalist's natural properties, could guide decision and action and thus satisfy (1), they would first have to engage or at least attract the will. Alternatively put, the folks' action-guiding subjective values would have to have been formed and conditioned as motivated responses to those natural properties that comprise the objective moral values. At least this is so if we believe, as moral naturalists do, that the central folk moral judgments are objectively true.

Even so, there is a puzzle with Mackie's subjective values that John Burgess has recently drawn to our attention (Burgess 1998). Mackie seems to have taken subjective values more or less for granted, apparently regarding the revelation that all our putative objective values are in reality subjective as a vindication of his Error

Theory. Yet, as Burgess observes, even if Mackie is correct in holding that the source of all our moral evaluations is to be found in our own attitudes rather than the objective natural facts, the fact is that we judge *actions* to be good or bad, right or wrong, and *people* to possess various moral attributes. If this is so, we are still committing the error of regarding as valuable something that is not in itself valuable at all. So why doesn't Mackie reject subjective values as well?

1.2 Moral Realism

We need to say a little more about how moral realism is to be construed. Without attempting a precise definition, I take moral realism to be the thesis that there are mind-independent moral entities and properties—norms, values, duties, and such like. Naturalistic moral realism is then the view that the mind-independent entities and properties that are the moral ones are simply certain sorts of natural entities or properties.

An ontological construal of moral realism is, in my view, preferable to any semantic construal since the latter are too weak to distinguish realist views from various antirealist pretenders. Thus, moral nihilists who are quietists about folk moral practice and deflationist about truth might agree with the claim that folk moral judgments are either true or false, on the grounds that it is part of the socially useful practice of issuing moral judgments or resolving ethical disputes that there be definite conditions under which we can assert some ethical claims and deny others.

Thomas Nagel defends an attractive form of moral realism that is grounded in the objectivity of moral reasoning. At first blush, Nagel's view looks as if it is not amenable to any type of ontological construal. For he writes:

I take it for granted that the objectivity of moral reasoning does not depend on its having an external reference. There is no moral analogue of the external world—a universe of moral facts that impinge upon us causally. Even if such a supposition made sense, it would not support the objectivity of moral reasoning (see p. 101 in Nagel 1997).

This sounds as if he may mean to reject real moral entities and properties as superfluous. He also seems to want to reject naturalistic moral realism if this means, as I have assumed it does, that moral entities and properties are reducible to natural entities and properties, when he claims: '... in the end, the contest is between the credibility of substantive ethics and the credibility of an external psychological reduction of that activity' (see p. 115 in Nagel 1997).

These appearances are however misleading. Nagel sees moral reasoning as a species of practical reasoning aimed not at finding out how things are in the world but rather at settling the question of what we ought to do. That is why the supposition that moral facts impinge upon us causally is superfluous even if coherent. Crudely, we act so as to change the world; it is irrelevant to that project that our sources of motivation stem from the world acting on us first (if indeed it does so).

As to moral naturalism, Nagel is perfectly happy to countenance the possibility that rule utilitarianism, for instance, might provide the best vindication of our moral

reasoning. The ‘external psychological reduction’ Nagel means to oppose is not a *bona fide* naturalistic reduction of value to some natural property but what he terms the subjectivist/relativist ‘reduction’ of objective moral reasoning to some merely parochial or subjective foible of our human psychological makeup. In fact, the latter ‘external psychological reduction’ is not a genuine *metaphysical* reduction at all but part of an error theory about the source of the illusion of objective moral reasoning.

Nagel’s project is then to show how ‘. . . attempts to get outside of the object language of practical reasons, good and bad, right and wrong . . . and see all such moral judgments as expressions of a contingent, non-objective perspective will eventually collapse before the force of the first-order judgments themselves’ (see p. 103 in Nagel 1997). As we have already noted, an ‘internalist’ project such as Nagel’s which seeks to vindicate our belief in objective norms and values by demonstrating that the practice of first-order moral reasoning would be incoherent in their absence does not solve the Hard Problem of naturalistic moral epistemology. Nor does it purport to. It does not show how the folk manage to target precisely those objective natural properties that are the moral properties as they make their moral evaluations. But if successful, it does comprise a crucial first step towards any such solution—at least we’d know the relevant natural properties are there in the world and that it is to those properties that our moral reasoning is *somehow* differentially sensitive.

Henceforth, the type of naturalistic moral realism that I shall examine is Nagelian realism. The Hard Problem will have to wait. After all, there is no certainty its solution even lies within our grasp. I am going to argue that whilst Nagel is right to contend that ‘the procedures of justification and criticism we employ can [not] be thought of as . . . something we just do,’ this does not by itself vindicate moral realism. There is an interesting type of moral agnosticism that concurs with Nagel’s appraisal.

1.3 Possible Conflict Between Moral Evaluations and Moral Theory

Nagel is not alone in counterposing first-order judgments to second-order judgments in order to confute skepticism. There is currently a lively dispute in philosophy of mind about the possibility of genuinely mental causation in a purely physical world.⁸ If mental properties are causally otiose, how is epiphenomenalism to be avoided? Philosophers such as Tyler Burge, Lynne Rudder Baker, and Robert van Gulick have appealed to the certainty we feel in our ordinary first-order ascriptions of mental causation as an antidote to the second-order philosophical scepticism engendered by naturalistic worries about how such causation is possible. Thus Burge claims:

Materialist metaphysics has been given more weight than it deserves. Reflection on our explanatory practice has been given too little. The metaphysical grounds that support the worries are vastly less strong than the more ordinary grounds we already have for rejecting them. . . . I think it more natural and fruitful to begin by assuming, defeasibly perhaps but firmly, that attributions of intentional mental events are central to psychological explanation

both in ordinary life and in various parts of psychology. We may also assume that intentional mental events are often causes and that psychological explanation is often a form of causal explanation. Given these assumptions, the ‘worry’ about epiphenomenalism seems very remote. . . . None of the metaphysical considerations advanced in current discussion seem to me remotely strong enough to threaten this conclusion (see p. 97 and p. 118 in Burge 1995).⁹

Jaegwon Kim responds to Burge as follows:

The problem of mental causation is primarily a metaphysical problem. It is the problem of showing how mental causation is possible not whether it is possible, although of course what happens with the how-question may in the end induce us to reconsider our stance on the whether-question. In raising the how-question, we are assuming “defeasibly but firmly” as Burge says, that the whether-question has already been affirmatively answered (see p. 61 in Kim 2000).

Kim’s reply is quite correct as far as it goes. However he only alludes to the underlying philosophical question that I think is really worrying Burge: *what if our theoretical deliberations lead us to conclude that mental causation is impossible yet we find ourselves unable as theorists to treat our own attributions of mental causation to ourselves and others as fictions or errors or mere rationalizations?* This is Nagel’s question and Burge tenders a Nagelian reply: in that event, we would have to disbelieve the theory since the first-order attributions of mental causation are much more secure epistemically than any of the various metaphysical theories about mental causation.

What of the moral case though? My judgment that incarceration of innocent children is objectively impermissible seems to me to be, if anything, even more compelling than my judgment that my desire for beer caused me to go to the refrigerator to look for it. I cannot see how I could possibly rescind the moral judgment even if, to my great surprise, it turned out that it was not my desire for beer that led me to the refrigerator. Moreover, like Nagel, I cannot see how I could regard it as just another fact about me that I happen to make this considered moral judgment rather than its contrary say or none at all.

The Hard Problem of moral epistemology is a metaphysical one also. It is the problem of showing how the folk detect and are differentially sensitive to those natural properties that are the moral ones. So, insofar as I am a naturalistic realist, must I follow Burge and Kim in assuming ‘defeasibly but firmly’ that the ‘whether question’ has been affirmatively answered in the moral case?

No. There is a way of respecting the perceived character of objective prescriptivity implicit in folk moral judgment that is compatible with agnosticism about the reality of the moral values and properties we take to underwrite them. In order to appreciate the type of moral agnosticism that can sustain this attitude, it is useful to view moral norms in a wider setting.

One important function of moral norms is to counteract limited sympathies, as Mackie put it, and if moral norms are perceived by the folk as objectively and intrinsically prescriptive, and *that they are so perceived* is common knowledge amongst them, then even if this perception is erroneous, these norms may function under favorable conditions as credible replacements for various forms of

external enforcement within social transactions that would otherwise yield Pareto sub-optimal outcomes.

The general observation is hardly new, comfortably predating Mackie. Indeed, it was clearly appreciated by Plato. Whilst the relativisation of the efficacy of norms to favorable conditions might sound a little unfamiliar in the moral context, only my proposal about the nature of those conditions is novel.

Game theorists distinguish, amongst agreements and threats, those that lack credibility from those that do not. Any agreement or threat that is not enforceable is deemed ‘cheap talk’. If moral norms are to function in the way they do to mediate folk social transactions, they *need* to be perceived precisely as the folk do perceive them: as intrinsically and objectively prescriptive. Any person who fails to observe the moral injunction, for example, to respect the basic rights of others, is ostracized in a number of ways since, according to folk moral thought, it is an intrinsic and objective fact about human beings that they have basic rights. ‘Respect the fundamental rights of others’ is an intrinsically and objectively prescriptive folk moral norm.

Suppose now that there is in reality nothing to this talk of basic rights, that the whole idea of humans as possessing such properties is a collectively beneficial anthropomorphic illusion. If this fact were to become common knowledge amongst the folk, it would undermine the role of moral norms in social life. But wouldn’t it equally do so for any theorist apprised of this discovery? The justification for fair treatment of others, including *inter alia* not incarcerating young children, could no longer be that humans have certain fundamental rights.

Now if Kant was right, our moral duties are the dictates of reason. A consequence of his view is that certain simple forms of the Prisoner’s Dilemma (PD) can in principle be dissolved—each autonomous agent in certain simple PD situations has a non-instrumentally rational reason to cooperate with the other, and that this is so is common knowledge amongst them since it is *a priori* deducible from a knowledge of the conditions of their own autonomy.

If moral duties are illusions though, not only are the simpler PD situations not (dis)soluble through *a priori* reasoning, but there is now a choice for the theorist where there was not before as to whether to abide by moral precepts at all for those interactive situations in which moral precepts normally mediate. The challenge to the reflective theorist will then be to find a reason for acting morally in those interactive situations where doing this goes against his own perceived best interests. If moral norms and values are illusions or social constructs or reflections of purely subjective preferences, then recognition of this fact must lead the theorist to reconsider their probative force. Moral norms would then stand in danger of becoming cheap talk for the theorist. Let us call this the *Ring of Gyges* objection since its upshot is that knowledge of the unreal status of the moral law is a ring that renders the cognoscenti invisible to the dictates of that law.

I think it is the *Ring of Gyges* objection that is really troubling Nagel in his discussions of moral subjectivism and relativism: if moral norms and values are not fully objective and real, then we have in the end no reason to act morally. Thus he writes, ‘The opposition here is between a theory about how things are and a practice that would be impossible if this was how things are’ (see p. 106 in Nagel 1997).

Even so, there is something very unsettling about Nagel's inference from our inability to *think* of our forms of reasoning as less than fully objective to their *actually being* fully objective. How can *our* inability to see the flaws in *our own* (practical reasoning) program possibly show that there are *no* such flaws? The inference looks like a simple *ad ignorantiam*. Still, Nagel is in good company in endorsing this inference. Frank Jackson has proffered a related response to Mackie's moral skepticism using *a priori* principles based upon interpretative constraints:

Any account of **M** (matured Folk Morality) which makes it all but impossible for any actions in the world as it is actually constituted to be right must be mistaken. It is a constraint on interpreting our moral discourse that some actions have the moral properties our discourse putatively ascribes to them (Jackson 1992).¹⁰

It is worth taking some time to explain why this particular *a priori* insight is flawed since, if it were right, the Hard Problem would simply dissipate. To begin, for realists, there is no direct inference from intelligibility constraints governing the interpretation of some discourse to a metaphysical conclusion concerning the nature of the world—that we perforce interpret moral discourse as quantifying over moral values in no way shows that there *are* moral values. At best, it shows that *we folk* cannot make sense of what our *fellow folk* are talking about unless we assume the existence of moral values. This says something about *us*, not the world.

Of course, one must observe the rules underwriting moral attribution to engage in the practice of moral evaluation and dispute. But this is perfectly compatible with retaining a healthy skepticism about the entities and properties the discursive practice apparently quantifies over. L.E.J. Brouwer, the founder of Intuitionism, furnishes a good example of this phenomenon—Brouwer made major contributions to classical mathematics, all the while maintaining that it was based on error.

If Jackson's idea is instead that since an interpreter must engage in the very practice she seeks to interpret, she must posit in the metalanguage the very same theoretical entities those she interprets in the object language, this is simply false. This would imply that Quine and the Churchlands are ineluctably committed to the existence of propositional attitudes by their understanding of the discourse that quantifies over those attitudes, that we could not understand primitive tribes unless we believed in their gods and so on.

Moreover, even if it were the case that interpreting propositional attitude discourse can only be achieved by attributing further propositional attitudes, or that interpreting moral attitude discourse could only be a matter of attributing moral attitudes, how would this provide an answer to the skeptic about either folk mentality or morality?

Consider a comparable case. Some logicians have argued that the *only* way to justify classical inference rules in the object language is to use classical reasoning in the metalanguage.¹¹ Clearly, this in no way shows that classical reasoning is itself justified in some absolute extrinsic sense. Indeed, opponents of classical logic such as Michael Dummett acknowledge this feature and use it as a ground of criticism—intuitionistic inference rules can be justified in a classical as well as intuitionistic metalanguage, whence the meanings of the intuitionistic logical constants

can be conveyed without circularity, or so some intuitionists hold.¹² Jackson seems to mistake the imposition of internal consistency requirements for the satisfaction of external correspondence relations.

Still, the idea that you can't interpret discourse D unless most of the D-terms and D-predicates refer to the things the D-users take themselves to be referring to, or more strongly, unless the things the D-terms refer to *have* the properties the D-predicates refer to, is one that seems to have some currency amongst philosophers. As far as I can see, there is no reason in the world to believe it. We have no difficulty interpreting the language of mediaeval cosmologists whose beliefs about the cosmos were widely false. Even in cases where the D-terms and D-predicates clearly do refer to their intended referents, interpretation can be successfully accomplished without imposing the extraneous requirement that the D-objects actually *have* the D-properties. Once objects have been assigned to terms and properties to predicates, semantic interpretation can be considered completed, at least for the purposes of understanding what D-speakers are talking about in the most obvious sense of 'talking about'.

At best, one might require, *a la* Quine (see Chapter 2 in Quine 1960) and Davidson, that unless the D-sentences describing simple and obvious matters turned out largely true *by the lights of the interpreter*, interpretation would be stymied in the sense that no accompanying propositional attitudes could plausibly be attributed to the D-users owing to the breach in rationality that would also have to be imputed to them in order to explain their cognitive failure. But both Quine and Davidson are quite happy to attribute systematic error to D-speakers, provided that it is rationally explicable from the interpreter's point of view.

The latter point highlights a crucial difference between Donald Davidson's argument against general skepticism and Jackson's related but more specific argument against moral skepticism, both being based on putative facts about interpretation. In an effort to secure the overall truth of our most basic beliefs about the world, Davidson appeals to the artifice of an omniscient interpreter constrained to use, as human interpreters perforce use, rationality constraints to determine which attitudes and meanings may justifiably be ascribed to the agent. Since the rationality constraints dictate that our basic beliefs turn out true by the interpreter's lights and since the interpreter in this case is omniscient, most of our basic beliefs *are* indeed true and the skeptic is stultified (Davidson 1984).

Absent the omniscient interpreter and Davidson has no reply at all to the skeptic. But Jackson makes no such appeal. What then secures the overall truth of the folk theory of morality? Why couldn't it be based on a systematic but seductive error just as Mackie argued?

1.4 Meta-Ethical Pyrrhonism

We are finally in a position to appreciate the type of moral agnosticism that can be used to undercut Nagel's argument for moral realism. I should first say that I do not see Nagel's argument for the objectivity of moral reasoning as in any way

fallacious. Rather than appealing as Jackson does to truths that can supposedly be known *a priori*, Nagel is offering an inference to the best explanation: what possible explanation could there be for our inability to think of our first-order moral evaluations as anything other than fully objective, than that they *are* fully objective?¹³ Like all inferences to the best explanation though, this argument fails if there is a better or at least equally plausible alternative explanation. In this case, I think there is.

The leading idea behind Meta-Ethical Pyrrhonism (**MEP**) (and the reason for the name) is that neither the available arguments *for* moral realism nor the arguments *against* are in the end rationally compelling—the strongest reason *for* thinking that there are real objective moral properties is that our practice of moral reasoning presupposes their existence and lapses into incoherence if there aren't any. The strongest argument *against* real moral properties is that no natural property could discharge the normative role required by our moral practices: there are no objectively and intrinsically prescriptive natural properties in the world. However, according to the Pyrrhonist, there is no breaking the dialectical stalemate since *neither set of judgments can be given up*.

Hence, like Pyrrhonism of yore, **MEP** holds that first-order folk moral judgments presupposing moral properties are real are exactly counterbalanced by naturalistic second-order judgments that moral properties are fictions/ constructs/illusions. The reason there can be no present resolution of the tension between first-order and second-order judgments is that whilst moral practice is indispensable to us, fictional moral properties seem unable to 'save all the moral phenomena': in particular *no ersatz moral property can give us a reason to act morally when it goes against our best interests to do so*. The upshot here is that for the theorist, interactive situations wherein moral norms can secure cooperation between folk in the teeth of mixed or conflicting interests now become potential PD situations or else situations where *the theorist* has no reason not to unilaterally defect, even if the other agent believes defection to be intrinsically and objectively wrong.¹⁴

MEP is thus a form of *meta-ethical agnosticism* that can nonetheless supply the theorist with a reason for acting morally in certain Pareto sub-optimal situations: According to the Pyrrhonist, we should suspend judgment on whether moral properties are real or not, but we should still act as if moral norms and laws were intrinsically and objectively prescriptive.

MEP is *not* moral skepticism at least if that is taken to mean that one cannot justify belief in any normative moral claim such as 'incarcerating children is wrong'. Neither is **MEP** the claim that *there is no fact to the matter* as to whether moral properties are real or that it is impossible to *discover* whether moral properties are real. **MEP** is instead the thesis that it is *as of now undecidable* whether moral properties are real or not. It is indexed to the information and theories that we currently have at our disposal (or foreseeable extensions thereof). Further information about the human brain or the evolution of cooperative behavior and reciprocity in hominids might supply essential clues that could be used to break the epistemic deadlock.

Despite being a form of moral agnosticism, **MEP** claims that we should acquiesce in the undeniable *appearances* of objective moral prescriptivity and judge and act in accord with folk morality. Yet we might well wonder *why* we should do this.

A religious agnostic has no reason to act/judge religiously, it might be argued, so why should a moral agnostic act/judge morally? The reason has to do with the very agnosticism that underwrites the Pyrrhonist's position: in a situation of radical uncertainty between two alternatives, if on either alternative one is better off in pursuing a certain course of action, then one should pursue that course of action. The relevant course of action is to obey the central moral norms the rest of the folk obey. If these are real and objective, one will have been saved from committing acts that are morally wrong; if they are not, one should still obey them since one is instrumentally better off in acting as if they are.¹⁵

Pyrrhonists believe that there is a 50% chance that there are objectively prescriptive moral norms and a 50% chance that the appearance of objective prescriptivity for moral norms is an evolutionarily useful illusion 'designed' to secure social cohesion. Then to 'cooperate' is to judge/act in accord with the precepts of Folk Morality. To defect is not to do so. The situation can be modeled game-theoretically as in Fig. 1.1 below with the agents radically uncertain as to whether the actual choice situation they face is a Prisoner's Dilemma (PD) or one wherein there are real moral norms. Roughly, if the second-order naturalistic arguments are correct, we are in a PD situation; if, on the other hand, folk first-order moral judgments are objectively true, we are in a moral realist (MR) situation.¹⁶

That Pyrrhonists are indeed better off by acting in accord with the dictates of folk morality can be seen from the final payoff matrix calculated for the situation of radical epistemic uncertainty in which they find themselves, as in Fig. 1.2. The cells of the matrix in Fig. 1.2 are calculated using a 50% contribution from each appropriate cell in the two matrices represented in Fig. 1.1. The solution to the game is then for each agent to cooperate by observing the norms of folk morality, joint cooperation comprising the Nash Equilibrium.¹⁷

An interesting feature of this game can be used to model the aspect of moral practice which so impressed Nagel: our inability to think of our moral judgments and decisions as anything other than fully objective. For whilst defection will reveal to an agent which situation s/he is in, defection is instrumentally irrational.

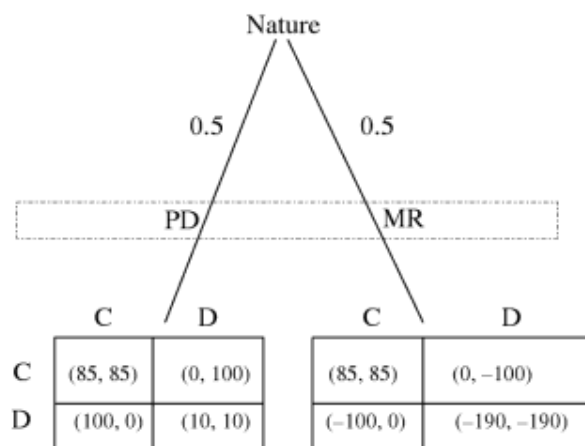


Fig. 1.1 MEP choice situation

Fig. 1.2 MEP payoff matrix

	Cooperate	Defect
Cooperate	(85, 85)	(0, 0)
Defect	(0, 0)	(-90, -90)

Cooperation through observance of the relevant moral norms, on the other hand, which is the rational course of action for each agent, will not reveal to the agents which situation they are actually in.

Those who cannot resist the temptation to indulge their epistemic curiosity might discover they are really in a PD situation. Nonetheless, it will be irrational to attempt to discover this fact in any situation where the assumption that moral norms exist works to the benefit of all agents. Hence, even if they are fictional constructs, moral norms can still bind the theorist so long as their fictional status remains hidden during moral transactions and cannot be subsequently demonstrated in reflection on the nature of those transactions. Thus is the *Ring of Gyges* objection disarmed by the Pyrrhonist.

It may be thought that even if there are sound instrumental reasons for cooperating with the dictates of folk morality, there can be no basis for a moral agnostic to endorse any moral judgments. Surely this, at least, is not in the spirit of the classical pyrrhonist who sought release from the vexatious tug of conflicting judgments by refusing to tender any?

The latter point is right in one sense. Meta-Ethical Pyrrhonists *do* differ from their classical forebears in issuing moral judgments and engaging in debate with a view to coming to some settled opinion about what ought to be done (rather than what *appears* as if it ought to be done)—that is, what first-order moral judgment or course of action one should endorse. They agree with their forebears, of course, in thinking that no settled second-order judgment can be reached as to the metaphysical nature of moral norms or values.

The justification for coming to an opinion about first-order judgments is one that classical Pyrrhonists did not consider, although Pascal most certainly did. Suppose with the Pyrrhonist that there is a 50% chance that moral norms and values are real and a 50% chance that they are instead evolutionarily useful illusions ‘selected for the purpose’ of counteracting limited sympathies amongst social animals. Suppose next that cooperation with those norms is beneficial in either case, as Pyrrhonists aver. Then it does not matter that the ‘ought’ as it features in the judgment ‘no child ought to be incarcerated’ is ambiguous between the deontic sense and the sense of practical rationality, the judgment comes out true on either reading.

So, finally, we have an answer to Nagel’s inference to the best explanation argument for moral realism. There is in fact a better explanation of our inability to think of our moral reasoning as anything other than objective: it is not in our best interests

to inquire about the reality of the putative objective norms that mediate various social transactions when we are engaged in these transactions, and since subsequent second-order naturalistic reflections on our first-order moral judgments are nowhere near robust enough to offset the hold of the latter on our decision-making or the benefits that accrue to acting upon those judgments, we should cooperate with the dictates of these moral precepts in spite of our uncertainty.

The effect will be the same: it will be irrational to treat moral norms in the context of decision-making as anything other than fully objective, even if it should prove theoretically impossible to subsequently think of those same norms as anything other than non-natural and thus non-objective properties from the naturalistic point of view.

1.5 Conclusion

My aim has been to undercut a persuasive argument for moral realism, not to promote an alternative that I confess to finding philosophically unsatisfying. For to the extent that one wishes to know what types of things the world contains, to the extent that one is interested in metaphysics in other words, one will not rest content with any form of moral agnosticism and will seek to advance new considerations or review old ones in an effort to decide whether moral properties are real or not.

However, not every philosopher shares a passion for metaphysics. One who most assuredly does not is Bas van Fraassen who professes to reject the very project of metaphysics on empiricist grounds.¹⁸ Someone sympathetic to van Fraassen's views might endorse Meta-Ethical Pyrrhonism as a stable and defensible position about the irrelevance of moral ontology to moral theorizing on the grounds that Moral Realism poses a pointless epistemic risk, following van Fraassen's argument for Constructive Empiricism:

If I believe a theory to be true and not just empirically adequate, my risk of being shown wrong is exactly the risk that the weaker entailed belief will conflict with actual experience. Meanwhile by avowing the stronger belief, I place myself in the position of being able to answer more questions . . . having a richer fuller picture of the world . . . a wealth of opinion I can dole out to those who wonder. But since the extra opinion is not additionally vulnerable, the risk is . . . illusory and so is the wealth. It is but empty strutting and posturing, this display of courage not under fire and avowal of additional resources that cannot feel the pinch of misfortune any sooner (see p. 255 in Churchland and Hooker (eds) 1985).

'Empirical adequacy' in the moral case will simply amount to a vindication of the central folk moral evaluations. So the suggestion is that a realist meta-ethics is a pointless epistemic risk caused by an unwelcome intrusion of metaphysics into a practice that is definitive of our humanity.

Those less hostile to metaphysics than van Fraassen might independently endorse Meta-Ethical Pyrrhonism as the only stable meta-ethics consistent with naturalism. For if it really is true, as Nagel argues, that we cannot surrender our first-order moral judgments or trade them in for subjectivist or relativist ersatzes, if we really are forced to regard our moral evaluations as intrinsically and objectively

prescriptive every time we issue them, and yet we can provide no subsequent naturalistic explanation of a property so strange as objective prescriptivity, the most plausible naturalistic response to this situation may well prove to be the Pyrrhonist's: *since our moral evaluations and our naturalistic reflections cannot be brought into harmony with each other and since neither set of judgments can be given up, as naturalists we have no choice but to remain agnostic about the reality of moral properties.*

Notes

¹ To the extent that they have not realized that the semantic attacks of antirealists on metaphysical realism are driven by a conviction that the Representation Problem is insoluble for realists. For the argument supporting this claim, see Khlentzos (2004).

² I have omitted from the quote the word van Fraassen is actually discussing which is the word 'part' but his point is equally applicable to the word 'wrong'.

³ For a compelling defense of the hypothesis that convulsive ergotism may have been the physiological basis for the Salem witchcraft crisis of 1692, see Capporeal (1976).

⁴ My conjecture is that fixation on the water-H₂O model has led to complacency amongst naturalists that the *Hard Problem* is no harder in the moral case than the water case. I shall explain why this model is a highly misleading one for a moral naturalist to appeal to.

⁵ Once more, though, no mere mapping between the two is sufficient to establish the identity of natural property with folk moral property.

⁶ Although only *prima facie* since the fact of moral disagreement may be attributable to disagreement about the non-moral facts.

⁷ For recent surveys of this evidence, see Bechara et al. (2000) and Dolan (2002). See Damasio (1994) for an accessible popular presentation.

⁸ For a representative selection of papers on this theme which includes papers from Kim, Burge, Rudder Baker and van Gulick, see Heil and Mele (eds) (1995).

⁹ Quoted on pp. 60–61 in Kim (2000).

¹⁰ In fact, although I think Jackson's argument is fallacious, I do not believe Nagel's is. Jackson's appeal to what can be known *a priori* is the difference between the two.

¹¹ Most notably Michael Dummett. See Dummett (1991).

¹² Cf. Dummett (1991).

¹³ *Qua* moral naturalist, Jackson is no more entitled to assume *a priori* that the extensions of 'right' and 'wrong' are non-empty than our fictitious theorist Nat is entitled to assume *a priori* that the extensions of 'cornucopian' and 'pandoran' are non-empty.

¹⁴ Indeed, as everyone knows, if the situation is a PD situation and the other agent is most likely to cooperate out of a sense of duty or a misplaced expectation of reciprocity or because s/he believes defection to be immoral, this supplies the canny theorist not bound by moral norms with the *best possible* reason to defect.

¹⁵ Readers will recognize the Pascalian character of this reasoning. Pascal, of course, famously denied that agnostics have no self-interested reason to act religiously.

¹⁶ Thanks to Arcady Blinov for making me aware of this choice-situation. See Blinov (2003) for a discussion of the use of such choice-situations in modeling the strategic benefits of what he sees as certain types of cognitive irrationality.

¹⁷ See Khlentzos (2004) for further elaboration and discussion of this choice-situation.

¹⁸ See his defense of this view in van Fraassen (2002).