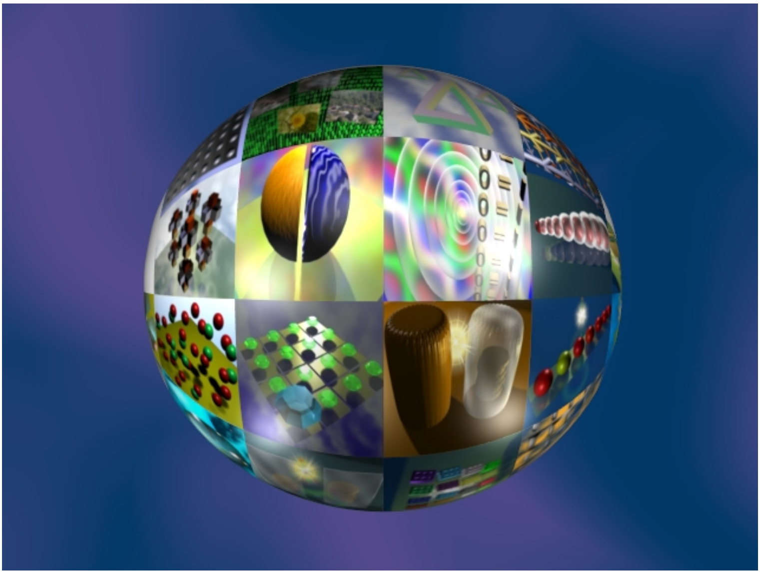
Paul Franceschi



# INTRODUCTION À LA PHILOSOPHIE ANALYTIQUE

## PARADOXES, ARGUMENTS ET PROBLÈMES CONTEMPORAINS

De P. à T.

Paul Franceschi

# INTRODUCTION À LA PHILOSOPHIE ANALYTIQUE

## PARADOXES, ARGUMENTS ET PROBLÈMES CONTEMPORAINS

Édition 2.1

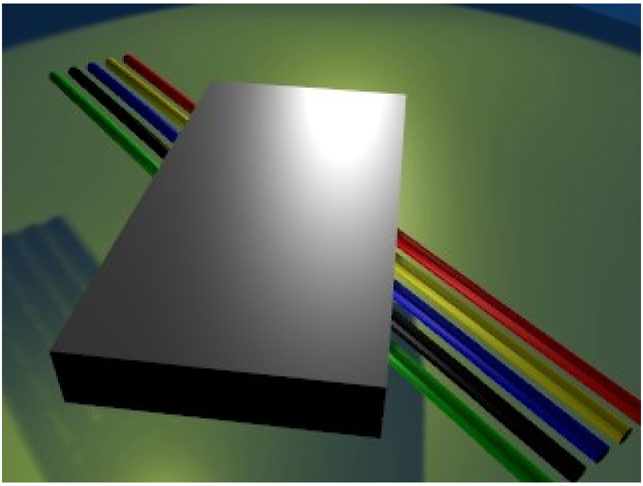


**Paul Franceschi**

Edition papier
TheBookEdition

**http://www.univ-corse.fr/~franceschi**

# Introduction



Le présent ouvrage se propose de constituer une introduction à la *philosophie analytique*. Il est essentiellement destiné au lecteur familiarisé avec la philosophie dite « continentale » et qui souhaite découvrir la philosophie analytique. Car ce style philosophique est souvent méconnu, en France notamment, où l'enseignement de la philosophie procède essentiellement d'une tradition nourrie par la philosophie « continentale ». Pour ma part, j'ai découvert pour la première fois les problèmes de philosophie analytique à travers les articles de Jean-Paul Delahaye publiés dans la revue *Pour la Science*. Je me souviens encore avec quel émerveillement j'ai découvert alors une façon d'appréhender la philosophie jusque là ignorée, qui correspondait à la tournure d'esprit qui, de manière naturelle, était la mienne. Si cette introduction, par bonheur, parvenait à faire partager au lecteur un peu de cet émerveillement, je crois qu'elle aurait alors atteint son but.

Le présent livre se propose ainsi de présenter un nombre significatif de problèmes contemporains en philosophie analytique. Il s'agit ici d'illustrer comment la démarche qui y est poursuivie consiste en la description précise de problèmes, clairement identifiés, et dont la présentation ne souffre pas d'ambiguïté. La démarche suivie tout au long de cet ouvrage consistera donc en la description d'un nombre important de problèmes philosophiques contemporains, illustrant ainsi la méthodologie utilisée en philosophie analytique qui consiste à décrire avec précision – souvent étape par étape – un certain nombre de problèmes bien identifiés, pour lesquels il n'existe pas, à l'heure actuelle, de solution consensuelle. Il apparaît utile, à cet effet, de classer les problèmes philosophiques contemporains en trois catégories distinctes : les paradoxes, les arguments et les problèmes proprement dits. Chacun de ces trois types de problèmes se trouve ici exposé, et accompagné le plus souvent d'une ou plusieurs solutions qui lui ont proposées dans la littérature contemporaine.

Je m'attacherai tout d'abord à décrire ainsi un certain nombre de paradoxes. Les plus célèbres d'entre eux trouvent leur origine dans l'Antiquité et ne sont toujours pas résolus : le Menteur, le paradoxe sorite, etc. Les paradoxes sont des arguments basés sur des prémisses et un raisonnement qui apparaissent tout à fait fondés, mais dont la conclusion conduit à une contradiction. Une excellente définition nous est fournie par Mark Sainsbury, dans son ouvrage *Paradoxes*, publié en 1995 : « Les paradoxes sont des conclusions inacceptables résultant d'arguments apparemment acceptables à partir de prémisses apparemment acceptables ».

Je présenterai ensuite un certain nombre d'arguments qui sont fréquemment débattus dans la littérature philosophique contemporaine. Le plus souvent, ces arguments constituent des raisonnements dont les prémisses et les déductions qui les accompagnent paraissent tout à fait acceptables, mais leur conclusion s'avère contraire à l'intuition. Les problèmes de ce type se distinguent des paradoxes en ce sens qu'ils ne conduisent pas véritablement à une contradiction. A la différence des paradoxes, on n'observe pas dans ce type d'arguments de contradiction proprement dite, mais seulement une conclusion qui se révèle contraire au bon sens et à l'ensemble de nos connaissances. Les arguments dont la conclusion se révèle contraire à l'intuition sont proches des paradoxes, en ce sens qu'il est très probable que le raisonnement qui les sous-tend soit fallacieux. En revanche, ces arguments se distinguent des paradoxes en ce sens que l'on ne peut écarter d'emblée la possibilité que notre intuition soit prise à défaut. Si tel était le cas, la solution apportée au problème posé par ce type d'argument se devrait alors d'expliquer pourquoi la conclusion en apparaît de prime abord contraire au bon sens.

Enfin, je décrirai un certain nombre de problèmes proprement dits qui ont donné lieu à des discussions récentes en philosophie analytique. Parmi ces problèmes basés sur des raisonnements, certains ont une origine très ancienne, alors que d'autres n'ont été décrits que très récemment.

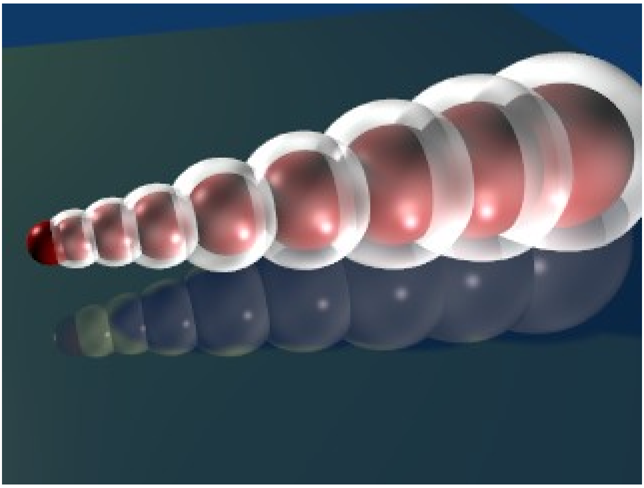
La philosophie analytique se caractérise essentiellement par une exigence de clarté dans l'exposition des idées et par un souci marqué de rigueur au stade de l'argumentation. La clarté des idées exprimées a pour but d'éviter l'ambiguïté et les difficultés liées à l'interprétation des textes. Elle permet également une meilleure évaluation critique des idées émises. Une telle exigence de rigueur nécessite parfois de faire appel à un formalisme mathématique, qui ne doit toutefois pas aller jusqu'à nécessiter des connaissances avancées en mathématiques. On le voit ici, la philosophie analytique constitue essentiellement un style philosophique.

Il est coutumier d'opposer la philosophie analytique et la philosophie continentale. La philosophie continentale se réfère ainsi aux écrits philosophiques d'auteurs français et allemands des XIXème et XXème siècles, parmi lesquels on peut citer – sans prétendre à l'exhaustivité : Friedrich Hegel, Sören Kierkegaard, Friedrich Nietzsche, Karl Marx, Herbert Marcuse, Martin Heidegger, Jean-Paul Sartre, Maurice Merleau-Ponty, Michel Foucault, etc. Les écrits de ces philosophes se caractérisent par une forme littéraire plus marquée et souvent un engagement politique plus poussé.

On associe parfois la philosophie analytique aux pays anglo-saxons et la philosophie continentale au continent européen. Un tel point de vue apparaît cependant assez réducteur. En effet, il est exact que la philosophie analytique constitue actuellement le style dominant au Royaume-Uni, aux États-Unis, au Canada, en Australie ou en Nouvelle-Zélande. Pourtant, elle s'avère également représentée en Europe, et notamment en France, en Italie, en Allemagne, en Espagne, au Portugal, en Grèce, en Belgique, etc. De plus, si l'on prend en considération l'antiquité et les philosophes classiques, il apparaît clairement qu'un tel point de vue se révèle erroné. Car on retrouvera notamment un style analytique très pur sur les bords de la Méditerranée, dans les écrits de plusieurs philosophes de l'antiquité. Les philosophes grecs classiques, inventeurs de paradoxes célèbres et non résolus tels que le paradoxe du Menteur, le paradoxe sorite, mais aussi les paradoxes de Zénon d'Elée, en constituent des exemples remarquables. Chez Platon également, on retrouvera aussi la clarté de l'argumentation dans la célèbre allégorie de la caverne. En outre, on trouvera chez Pascal, avec l'argument du pari, tous les critères d'une argumentation détaillée, précise et claire, qui satisfait tous les canons de la philosophie analytique contemporaine. Et surtout, on pourra constater que Descartes pratiquait avant l'heure un style analytique étonnamment pur. Nombre des arguments de Descartes auraient pu figurer sans changement dans la littérature analytique contemporaine. Dans le présent ouvrage, on trouvera ainsi le célèbre argument du cogito, l'argument du mauvais génie, l'argument ontologique de Descartes, ainsi qu'un argument en faveur du dualisme corps/esprit.

Il serait plutôt maladroit et manichéen d'opposer les deux styles – analytique et continental – en considérant que l'un est meilleur que l'autre. De manière moins ouvertement subjective, on peut estimer qu'il s'agit là de deux styles différents de pratiquer la philosophie, qui possèdent chacun leurs avantages et leurs inconvénients. Il apparaît très certainement nécessaire de préserver à la fois l'un et l'autre, compte tenu de leurs mérites respectifs et de leur complémentarité. Finalement, il apparaît que la coexistence des deux styles constitue essentiellement l'expression d'une diversité culturelle qui se révèle elle-même synonyme de richesse.

# 1. Le paradoxe du Menteur



Le paradoxe du Menteur constitue l'un des plus anciens et des plus profonds paradoxes connus. Il est attribué au philosophe grec Eubulide de Milet, qui vivait au IVème siècle avant J-C. Le paradoxe du Menteur peut être exprimé très simplement, car il naît directement de la prise en compte de l'affirmation suivante : « Cette phrase est fausse ». Le paradoxe provient du fait que si cette dernière phrase est *vraie*, alors il s'ensuit qu'elle est fausse ; mais si cette même phrase est *fausse*, alors il est faux qu'elle est fausse et donc qu'elle est vraie. Ainsi « Cette phrase est fausse » est fausse si elle est vraie, et vraie si elle est fausse. En conclusion, « Cette phrase est fausse » est vraie si et seulement si elle est fausse. Et cette dernière conclusion se révèle paradoxale.

On dénote souvent « Cette phrase est fausse » par ($\lambda$). Il est utile à ce stade, de décrire de manière détaillée les différentes étapes du raisonnement qui conduisent au paradoxe du Menteur (le symbole $\therefore$ dénote ici la conclusion) :

| | | |
|---|---|---|
| $(\lambda)$ | $(\lambda)$ est *fausse* | |
| (1) | $(\lambda)$ est soit *vraie* soit *fausse* | *bivalence* |
| (2) | si $(\lambda)$ est *vraie* | *hypothèse 1* |
| (3) | alors il est *vrai* que $(\lambda)$ est *fausse* | *de* $(\lambda)$*,(2)* |
| (4) | alors $(\lambda)$ est *fausse* | *de (3)* |
| (5) | si $(\lambda)$ est *fausse* | *hypothèse 2* |
| (6) | alors il est *faux* que $(\lambda)$ est *fausse* | *de* $(\lambda)$*,(5)* |
| (7) | alors $(\lambda)$ est *vraie* | *de (6)* |
| (8) | $\therefore$ $(\lambda)$ n'est ni *vraie* ni *fausse* | *de (4),(7)* |

La conclusion (8) est ici paradoxale, car il s'ensuit que $(\lambda)$ n'est ni *vraie* ni *fausse*, en contradiction avec le principe (1) de *bivalence*. Le problème que soulève le Menteur est ainsi le suivant : quelle est donc la *valeur de vérité* de la proposition $(\lambda)$, étant donné qu'on ne peut lui attribuer, sans contradiction, la valeur de vérité *vrai* ou *faux* ?

Une première tentative de solution pour le Menteur consiste à considérer que la valeur de vérité de $(\lambda)$ n'est ni *vrai* ni *faux*, mais une troisième valeur de vérité : *indéterminé*. On considère ainsi une logique tri-valuée, qui comporte ainsi les trois valeurs de vérité suivantes : *vrai*, *faux*, *indéterminé*. Le Menteur se trouve alors réintroduit sous la forme suivante :

$(\lambda_3)$ $(\lambda_3)$ est *fausse* ou *indéterminée*

Dans ce nouveau contexte, une proposition peut désormais prendre trois valeurs de vérité différentes : *vrai*, *faux* ou *indéterminé*. Le principe de tri-valence stipule alors que $(\lambda_3)$ est soit *vraie*, soit *fausse*, soit *indéterminée*. Cependant, le fait de considérer tour à tour que $(\lambda_3)$ est *vraie*, *fausse*, ou bien *indéterminée* ne conduit toujours pas à une solution satisfaisante, car il s'ensuit, en vertu du même raisonnement qu'avec le Menteur simple, la conclusion selon laquelle $(\lambda_3)$ n'est ni *vraie*, ni *fausse*, ni *indéterminée*. Il en résulte ainsi l'impossibilité d'assigner valablement une valeur de vérité à la proposition $(\lambda_3)$.

Plus encore, il apparaît que le problème resurgit de la même manière si on considère non plus trois, mais quatre valeurs de vérité : *vrai*, *faux*, *indéterminé*$_1$ et *indéterminé*$_2$. On utilise alors une logique 4-valuée. Cependant, il en résulte la variation suivante du Menteur :

($\lambda_4$) ($\lambda_4$) est *fausse* ou *indéterminé*$_1$ ou *indéterminé*$_2$

qui conduit de même que précédemment à l'impossibilité d'attribuer une valeur de vérité à ($\lambda_4$).

Une autre tentative de solution consiste alors à rejeter le principe de bivalence, de tri-valence, et plus généralement de *n*-valence sur lequel est basé le raisonnement auquel conduit le Menteur. Cependant, une telle tentative de solution échoue également, car elle se heurte à une variation plus puissante encore du Menteur, le *Menteur renforcé*, qui ne nécessite pas de faire appel à un quelconque principe de bivalence, de 3-valence, ..., ou de *n*-valence :

($\lambda_s$) ($\lambda_s$) est non-*vraie*

Car le Menteur renforcé conduit au raisonnement suivant :

| | | |
|---|---|---|
| ($\lambda_s$) | ($\lambda_s$) est non-*vrai* | |
| (9) | ($\lambda_s$) est soit *vrai* soit *non-vrai* | *dichotomie* |
| (10) | si ($\lambda_s$) est *vrai* | *hypothèse 1* |
| (11) | alors il est *vrai* que ($\lambda_s$) est non-*vrai* | *de ($\lambda_s$),(10)* |
| (12) | alors ($\lambda_s$) est non-*vrai* | *de (11)* |
| (13) | si ($\lambda_s$) est *non-vrai* | *hypothèse 2* |
| (14) | alors il est non-*vrai* que ($\lambda_s$) est non-*vrai* | *de ($\lambda_s$),(13)* |
| (15) | alors ($\lambda_s$) est *vrai* | *de (14)* |
| (16) | ∴ ($\lambda_s$) n'est ni *vrai* ni non-*vrai* | *de (12),(15)* |

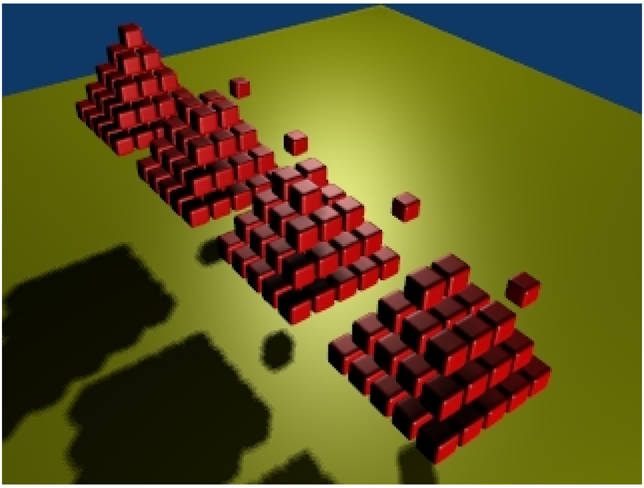
Enfin, une autre tentative de solution pour le paradoxe du Menteur consiste à considérer que la structure du

Menteur est auto-référentielle, puisqu'une telle proposition fait directement référence à elle-même. Selon ce type de solution, il suffirait d'interdire la formation des propositions auto-référentielles pour empêcher l'apparition du paradoxe. Cependant, une telle solution apparaît trop restrictive, car il existe de nombreuses propositions dont la structure est auto-référentielle, mais pour lesquelles l'attribution d'une valeur de vérité ne pose aucun problème. Il suffit de considérer pour cela le *Menteur contingent* :

($\lambda_c$) soit cette proposition est fausse, soit 0 = 0

Or il s'avère que l'on peut attribuer valablement la valeur de vérité *vrai* au Menteur contingent. Ainsi, bien que le Menteur contingent présente une structure auto-référentielle, on peut lui attribuer sans contradiction, à la différence du Menteur, une valeur de vérité. Dans ce contexte, il apparaît que le fait de proscrire purement et simplement toutes les propositions auto-référentielles conduirait à payer un prix trop élevé pour résoudre le paradoxe du Menteur, et ne constitue donc pas non plus une solution satisfaisante.

## 2. Le paradoxe sorite



Le paradoxe sorite (ou paradoxe du tas) est un des plus anciens et des plus importants paradoxes connus. On attribue son origine à Eubulide de Milet, le philosophe grec de l'antiquité auquel on doit également le paradoxe du Menteur. Le paradoxe peut être décrit, de manière informelle, de la façon suivante. Il est tout d'abord communément admis qu'un ensemble comportant 100000 grains de sable est un tas. De plus, il apparaît que si un ensemble comportant un nombre donné de grains de sable est un tas, alors un ensemble comportant un grain de sable de moins est également un tas. Compte tenu de ces prémisses, il s'ensuit la conclusion selon laquelle un ensemble comportant un seul grain de sable est également un tas. En effet, si un ensemble comportant 100000 grains de sable est un tas, il s'ensuit qu'un ensemble comportant 99999 grains de sable est un tas ; et il en va de même pour un ensemble comportant 99998 grains de sable, puis 99997, 99996, 99995, ..., et ainsi de suite, jusqu'à un seul grain de sable. Le paradoxe provient du fait que le raisonnement correspondant apparaît tout à fait valide, alors que la conclusion qui en découle se révèle inacceptable.

Les différentes étapes qui conduisent au paradoxe sorite peuvent détaillées de la manière suivante :

(1) un ensemble comportant 100000 grains de sable est un tas
(2) si un ensemble comportant $n$ grains de sable est un tas, alors un ensemble comportant $n$ - 1 grains de sable est un tas
(3) si un ensemble comportant 100000 grains de sable est un tas, alors un ensemble comportant 99999 grains de sable est un tas
(4) ∴ un ensemble comportant 99999 grains de sable est un tas
(5) si un ensemble comportant 99999 grains de sable est un tas, alors un ensemble comportant 99998 grains de sable est un tas
(6) ∴ un ensemble comportant 99998 grains de sable est un tas
(7) si un ensemble comportant 99998 grains de sable est un tas, alors un ensemble comportant 99997 grains de sable est un tas
(8) ∴ un ensemble comportant 99997 grains de sable est un tas
(9) ...
(10) ∴ un ensemble comportant 1 grain de sable est un tas

La conclusion du paradoxe résulte de l'utilisation répétée d'un principe logique communément admis qui est dénommé *modus ponens*, et qui présente la forme suivante : $p$, si $p$ alors $q$, donc $q$ (où $p$ et $q$ dénotent deux propositions).

On rencontre dans la littérature de nombreuses variations du paradoxe sorite. Une autre version du paradoxe avec le prédicat *grand* est ainsi la suivante :

| | | |
|---|---|---|
| (11) | un homme qui mesure 200 cm est grand | *prémisse de base* |
| (12) | si un homme qui mesure *n* cm est grand, alors un homme qui mesure *n* - 1 cm est grand | *prémisse d'induction* |
| (13) | ... | |
| (14) | ∴ un homme qui mesure 140 cm est grand | |

De même, on peut également construire des variations du paradoxe avec d'autres concepts vagues tels que *riche*, *vieux*, *rouge*, etc. Ceci conduit à mettre ainsi en évidence la structure suivante du paradoxe (où P dénote un prédicat vague) :

| | | |
|---|---|---|
| (15) | P(100000) | *prémisse de base* |
| (16) | si P(*n*) alors P(*n* - 1) | *prémisse d'induction* |
| (17) | ... | |
| (18) | ∴ P(1) | |

On peut observer ici que la structure du paradoxe est réversible. En effet, les versions précédentes du paradoxe procèdent par décrémentation. Mais le paradoxe peut également opérer par incrémentation, de la manière suivante :

| | | |
|---|---|---|
| (19) | un homme qui possède 1 cheveu est chauve | *prémisse de base* |
| (20) | si un homme qui possède *n* cheveux est chauve, alors un homme qui possède *n* + 1 cheveux est chauve | *prémisse d'induction* |

(21) ...
(22) ∴ un homme qui possède 100000 cheveux est chauve

La structure du paradoxe est alors la suivante (P dénotant un prédicat vague) :

| | | |
|---|---|---|
| (23) | P(1) | *prémisse de base* |
| (24) | si P($n$) alors P($n$ + 1) | *prémisse d'induction* |
| (25) | ... | |
| (26) | ∴ P(100000) | |

De nombreuses solutions ont été proposées pour résoudre le paradoxe sorite. Cependant, aucune d'entre elles ne s'est révélée jusqu'à présent satisfaisante. Ainsi, le paradoxe sorite demeure toujours l'un des paradoxes contemporains les plus étudiés.

Une solution qui met en cause l'étape d'induction a notamment été proposée pour résoudre le paradoxe. Un tel type de solution est basé sur une approche par degrés et fait valoir ainsi que l'étape d'induction n'est vraie que pour certaines instances – les instances propres – de la notion de tas. Une telle analyse repose sur le fait que la notion de tas constitue une notion *vague*. Une notion de ce type se caractérise ainsi par l'existence d'instances propres (par exemple une valeur de $n$ égale à 1000000), de contre-instances propres (par exemple une valeur de $n$ égale à 2), mais aussi de cas-limites (par exemple une valeur de $n$ égale à 100) qui constituent une zone de pénombre entre les notions de tas et de non-tas. Selon l'approche par degrés, la valeur de vérité de l'étape d'induction est 1 lorsqu'on est en présence d'instances propres. Mais lorsqu'il s'agit de cas-limites, sa valeur de vérité est inférieure à 1. Il s'ensuit finalement que la valeur de vérité de l'étape d'induction, lorsqu'on prend en compte toutes les valeurs possibles de *n*, est légèrement inférieure à 1. Et ceci suffit à bloquer partiellement le processus déductif et à empêcher de parvenir finalement à la conclusion finale.

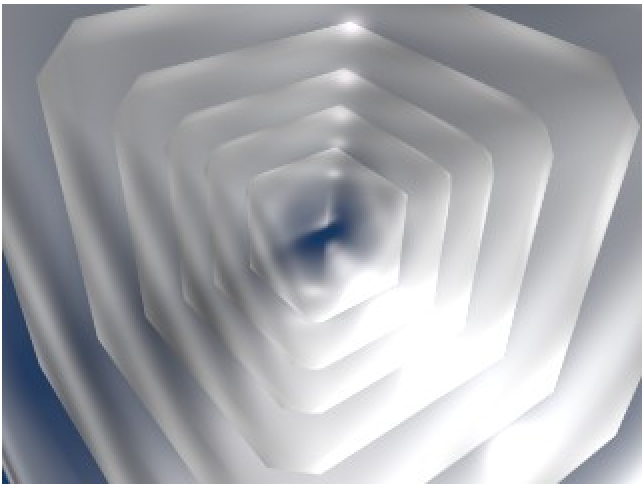
L'étape d'induction est également visée dans un autre type de solution qui considère que l'étape d'induction n'est pas nécessairement vraie. Il suffit par exemple de considérer une pile constituée de cubes empilés les uns sur les autres. Une telle pile peut comporter par exemple jusqu'à 20 cubes empilés. Maintenant, le raisonnement qui conduit au paradoxe sorite peut également s'appliquer à cette pile, car intuitivement, si on enlève les cubes un par un à partir du haut, on se trouve toujours en présence d'une pile. Pourtant, en réalité, on ne peut enlever certains cubes d'importance stratégique sans que tous les autres ne tombent d'un seul coup en détruisant en même temps l'ensemble de la pile. A l'inverse, certains cubes – notamment ceux du dessus – apparaissent moins stratégiques, de sorte qu'on peut les enlever sans compromettre l'existence même de la pile. Une telle analyse du paradoxe sorite suggère qu'il existe d'autres facteurs qu'il convient de prendre en compte tels que la position de chacun des cubes, leur alignement, etc. Cependant, un tel type de solution échoue également, car il se heurte à une variation purement numérique du même problème qui constitue le paradoxe de Wang :

| | | |
|---|---|---|
| (27) | 100000000 est *grand* | *prémisse de base* |
| (28) | si *n* est *grand* alors *n* - 1 est *grand* | *prémisse d'induction* |
| (29) | ... | |
| (30) | ∴ 1 est *grand* | |

En effet, un tel problème constitue une instance du paradoxe sorite, pour laquelle le type de solution précédent ne trouve désormais plus à s'appliquer.

Enfin, selon une autre approche, de nature épistémologique, il existe véritablement une frontière précise au niveau du nombre de grains permettant de différencier un tas d'un non-tas, mais il ne nous est pas possible de connaître précisément où se situe une telle frontière. La cause du paradoxe réside donc dans une déficience au niveau de nos connaissances, qui constitue ainsi une sorte de zone aveugle. Une telle frontière précise existe également, selon ce type d'approche, au niveau des notions de jeune/non-jeune, petit/non-petit, chauve/non-chauve, etc., en permettant ainsi de les distinguer. On le voit, un tel type de solution tend à rejeter l'étape d'induction comme fausse. Cependant, une telle solution ne se révèle pas non plus satisfaisante, car l'existence pour chaque notion vague, d'une coupure numérique précise permettant de distinguer les instances des contre-instances propres, apparaît plutôt contraire à l'intuition. Et un tel type de solution ne permet pas de rendre justice à l'intuition selon laquelle il existe, pour chaque concept vague, une zone de pénombre correspondant à des cas-limites.

## 3. Le paradoxe de Russell



Le paradoxe de Russell constitue l'un des paradoxes les plus fameux de la théorie mathématique des ensembles. Le paradoxe, énoncé par Bertrand Russell résulte, de manière informelle, de la prise en considération de l'ensemble de tous les ensembles qui ne se contiennent pas eux-mêmes. L'existence même de cet ensemble conduit directement à une contradiction. En effet, il s'ensuit d'une part que si cet ensemble appartient à lui-même, alors il n'appartient pas à lui-même. Et d'autre part, s'il n'appartient pas à lui-même, alors il appartient à lui-même. Ainsi, un tel ensemble, à la fois n'appartient pas à lui-même et appartient à lui-même.

Une variation classique du paradoxe de Russell est le problème du barbier. Un tel barbier rase tous les hommes qui ne se rasent pas eux-mêmes, et seulement ceux-là. La question qui s'ensuit est la suivante : ce barbier se rase-t-il lui-même ? Si le barbier se rase lui-même, alors par définition, il appartient à la classe des hommes qui se rasent eux-mêmes et par conséquent, il ne se rase pas lui-même. En revanche, si le barbier ne se rase pas lui-même, alors par définition, il appartient alors à la classe des hommes qui ne se rasent pas eux-mêmes et par conséquent, il se rase lui-même. En conclusion, si le barbier se rase lui-même, alors il ne se rase pas lui-même ; et s'il ne se rase pas lui-même, alors il se rase lui-même. Ainsi, que l'on considère l'une ou l'autre des hypothèses, il s'ensuit une contradiction.

Une autre version du paradoxe de Russell se présente sous la forme suivante : on considère le catalogue de tous les catalogues qui ne se mentionnent pas eux-mêmes. Il s'ensuit la question suivante : ce catalogue se mentionne-t-il lui-même ? S'il se mentionne lui-même, alors il ne fait pas partie de ce catalogue et ne se mentionne donc pas lui-même ; et s'il ne se mentionne pas lui-même, alors il fait partie du catalogue et se mentionne donc lui-même. Dans les deux cas, on se trouve en présence d'une contradiction.

Le paradoxe de Russell peut être énoncé ainsi de manière plus formelle. Soit R l'ensemble de tous les ensembles qui ne se contiennent pas eux-mêmes. On a ainsi la définition suivante de R (où $\in$ dénote l'appartenance à un ensemble et $\notin$ la non appartenance) :

(1) $x \in \mathrm{R} \mid x \notin x$

Maintenant, compte tenu de cette définition générale, on considère le cas particulier de l'ensemble R. Deux cas sont possibles : soit R appartient à lui-même, soit R n'appartient pas à lui-même. Dans l'hypothèse où R appartient à lui-même, le raisonnement s'établit comme suit :

| | | |
|---|---|---|
| (2) | $\mathrm{R} \in \mathrm{R}$ | hypothèse 1 |
| (3) | $\mathrm{R} \notin \mathrm{R}$ | de (2) |

Et de même, dans l'hypothèse où R n'appartient pas à lui-même, il s'ensuit, par définition :

| | | |
|---|---|---|
| (4) | $R \notin R$ | hypothèse 2 |
| (5) | $R \in R$ | de (4) |

La conclusion qui en résulte est que l'ensemble R appartient à lui-même si et seulement s'il n'appartient pas à lui-même. Les différentes étapes du raisonnement peuvent ainsi être détaillées :

| | | |
|---|---|---|
| (6) | $x \in R \mid x \notin x$ | *définition* |
| (7) | $R \in R$ | *hypothèse 1* |
| (8) | $R \notin R$ | *de (6),(7)* |
| (9) | $\therefore$ si $(R \in R)$ alors $(R \notin R)$ | *de (7),(8)* |
| (10) | $R \notin R$ | *hypothèse 2* |
| (11) | $R \in R$ | *de (6),(10)* |
| (12) | $\therefore$ si $(R \notin R)$ alors $(R \in R)$ | *de (10),(11)* |
| (13) | $\therefore R \notin R$ et $R \in R$ | *de (9),(12)* |

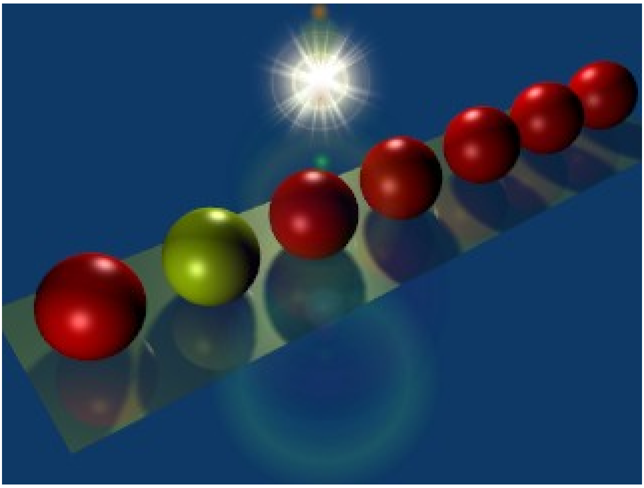
Ainsi, la prise en compte de l'existence de l'ensemble R de tous les ensembles qui ne se contiennent pas eux-mêmes conduit directement à une contradiction.

Le paradoxe trouve son origine dans la théorie naïve des ensembles, dans laquelle il est permis de définir un ensemble sans restriction. La théorie naïve des ensembles s'avérait ainsi trop libérale, en autorisant la construction de certains ensembles dont la nature se révélait finalement contradictoire, tels que l'ensemble R. En particulier, il est apparu que *l'axiome de compréhension* de la théorie naïve des ensembles se trouvait à l'origine de l'émergence du paradoxe de Russell. L'axiome de compréhension permettait en effet la construction de tout ensemble qui répondait au schéma suivant :

(14) $\mathrm{x} \in E \mid P(x)$

où P($x$) dénote une propriété quelconque présentée par un objet $x$, de sorte que tout $x$ présentant la propriété P appartient à l'ensemble E. Aussi, la solution pour résoudre le paradoxe de Russell, a-t-elle consisté à restreindre le pouvoir d'expression de la théorie des ensembles. Les axiomes de la théorie des ensembles ont ainsi été modifiés de manière à rendre impossible la construction de l'ensemble R de tous les ensembles qui ne se contiennent pas eux-mêmes. En 1908, Ernst Zermelo proposa ainsi une théorie des ensembles comportant un axiome de compréhension modifié, qui ne permettait plus la construction de l'ensemble R. Il en est résulté la théorie des ensembles de Zermelo-Fraenkel, qui est toujours en vigueur actuellement, et dont les axiomes rendent impossible la construction de l'ensemble R, évitant ainsi la contradiction qui en résulte.

# 4. Le paradoxe de l'examen-surprise



Le paradoxe de l'examen-surprise trouve son origine, dit-on, dans une annonce faite par les autorités suédoises durant la dernière guerre mondiale. Selon cette annonce, un exercice de défense civile était programmé pour la semaine suivante, sans que le jour précis n'en soit toutefois révélé, afin que l'exercice ait véritablement lieu par surprise. Le professeur Lennart Elkbom comprit le problème subtil qui résultait de cette annonce et en fit part à ses étudiants. Par la suite, le problème se répandit dans les cercles universitaires et donna ensuite lieu à de nombreuses discussions.

Le paradoxe de l'examen-surprise est classiquement décrit de la manière suivante. Un professeur annonce à ses étudiants qu'un examen aura lieu la semaine prochaine. Cependant, le professeur ajoute qu'il ne sera pas possible aux étudiants de connaître à l'avance la date de l'examen, car celui-ci aura lieu par surprise. Un étudiant intelligent raisonne alors ainsi : l'examen ne peut se dérouler le dernier jour de la semaine – vendredi – car sinon je saurai, de manière certaine, que l'examen aura lieu le vendredi. Ainsi, le vendredi peut-il être éliminé. De même, poursuit l'étudiant, l'examen ne peut se dérouler l'avant-dernier jour de la semaine – jeudi – car sinon je saurai que l'examen aura lieu le jeudi. Ainsi, le jeudi est-il également éliminé. Par le même raisonnement, l'étudiant conclut que l'examen ne peut avoir lieu ni le mercredi, ni le mardi, ni le lundi. Finalement, l'étudiant conclut que l'examen ne peut avoir lieu aucun jour de la semaine. Pourtant, cela n'empêche pas l'examen d'avoir lieu par surprise, par exemple le mercredi. Le paradoxe naît ici du fait que le raisonnement de l'étudiant semble valide, alors qu'il se révèle finalement en contradiction avec les faits, puisque l'examen a finalement bien lieu par surprise.

Le raisonnement de l'étudiant qui conduit au paradoxe de l'examen-surprise peut être détaillé de la manière suivante :

| | | |
|---|---|---|
| (1) | si l'examen a lieu le vendredi | *hypothèse 1* |
| (2) | alors je saurai que l'examen aura lieu le vendredi | *de (1)* |
| (3) | alors l'examen n'aura pas lieu par surprise | *de (2)* |
| (4) | ∴ l'examen ne peut avoir lieu le vendredi | *de (1),(3)* |
| (5) | si l'examen a lieu le jeudi | *hypothèse 2* |
| (6) | alors je saurai que l'examen aura lieu le jeudi | *de (5)* |
| (7) | alors l'examen n'aura pas lieu par surprise | *de (6)* |
| (8) | ∴ l'examen ne peut avoir lieu le jeudi | *de (5),(7)* |
| (9) | si l'examen a lieu le mercredi | *hypothèse 3* |
| (10) | alors je saurai que l'examen aura lieu le mercredi | *de (9)* |
| (11) | alors l'examen n'aura pas lieu par surprise | *de (10)* |

| | | |
|---|---|---|
| (12) | ∴ l'examen ne peut avoir lieu le mercredi | *de (9),(11)* |
| (13) | si l'examen a lieu le mardi | *hypothèse 4* |
| (14) | alors je saurai que l'examen aura lieu le mardi | *de (13)* |
| (15) | alors l'examen n'aura pas lieu par surprise | *de (14)* |
| (16) | ∴ l'examen ne peut avoir lieu le mardi | *de (13),(15)* |
| (17) | si l'examen a lieu le lundi | *hypothèse 5* |
| (18) | alors je saurai que l'examen aura lieu le lundi | *de (17)* |
| (19) | alors l'examen n'aura pas lieu par surprise | *de (18)* |
| (20) | ∴ l'examen ne peut avoir lieu le lundi | *de (17),(19)* |
| (21) | ∴ l'examen ne peut avoir lieu aucun jour de la semaine | *de (4),(8), (12),(16), (20)* |

Plusieurs solutions ont été proposées pour résoudre le paradoxe de l'examen-surprise. Aucune d'entre elles ne fait toutefois actuellement l'objet d'un consensus. Une première tentative de solution est apparue avec O' Connor, dans un article paru en 1948 dans la revue *Mind*. Selon lui, le paradoxe est dû au caractère contradictoire qui résulte de l'annonce du professeur et de sa mise en oeuvre. Pour O' Connor, l'annonce du professeur selon laquelle l'examen doit survenir par surprise se trouve en contradiction avec les données connues de la mise en oeuvre de l'examen. Ainsi, l'énoncé du paradoxe de l'examen-surprise est-il, selon O' Connor, auto-réfutant. Cependant, une telle analyse ne s'est pas avérée satisfaisante, car il est apparu que l'examen pouvait finalement survenir par surprise, sans contradiction, par exemple le mercredi. Et le fait que l'examen puisse en définitive survenir par surprise, confirmait bien l'annonce du professeur, sans la réfuter.

Une second type de solution a également été proposé par Quine, qui a mis en évidence le fait que quatre possibilités se présentent (en dénotant le dernier jour de la semaine par *n*) :

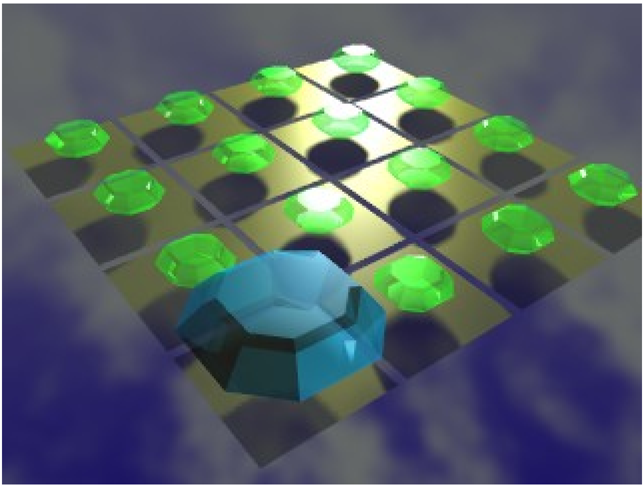
(a) l'examen aura lieu le jour *n* et l'étudiant saura que l'examen aura lieu le jour *n*
(b) l'examen aura lieu le jour *n* et l'étudiant saura que l'examen n'aura pas lieu le jour *n*
(c) l'examen n'aura pas lieu le jour *n* et l'étudiant saura que l'examen aura lieu le jour *n*
(d) l'examen n'aura pas lieu le jour *n* et l'étudiant saura que l'examen n'aura pas lieu le jour *n*

Selon Quine, le problème provient du fait que l'étudiant, au moment où il établit son raisonnement, n'envisage que les cas de figure (a) et (d), sans tenir compte des possibilités (b) et (c). En particulier, il ne prend pas en considération le cas de figure (b) qui est la situation réelle dans lequel il se trouve finalement, en permettant ainsi à l'examen de se dérouler finalement par surprise. Mais si l'étudiant avait envisagé cette possibilité dès le début, conclut Quine, il ne serait pas parvenu à une conclusion erronée.

Au titre des solutions, il a également été proposé que le paradoxe de l'examen-surprise se réduit au paradoxe sorite. Un tel point de vue a notamment été exposé, avec des nuances différentes par P. Dietl en 1973 et J. W. Smith en 1984. Ces deux auteurs font valoir que les deux paradoxes présentent une structure commune, de sorte que le paradoxe de l'examen-surprise se révèle finalement équivalent au paradoxe sorite. Selon une telle analyse, les différentes étapes des deux paradoxes sont équivalentes et le paradoxe de l'examen-surprise trouve ainsi son origine dans le fait que la surprise constitue une notion vague.

Mais une telle analyse a toutefois été critiquée par Roy Sorensen, dans son ouvrage *Blindspots*, publié en 1988, où il fait valoir que les deux problèmes ne sont pas réellement de même nature. En premier lieu, fait en effet valoir Sorensen, la version du paradoxe sorite équivalant au paradoxe de l'examen-surprise serait bien trop rapide. Et en second lieu, ajoute Sorensen, les prémisses de base des deux paradoxes ne peuvent pas véritablement être considérées comme équivalentes.

# 5. Le paradoxe de Goodman



Le paradoxe de Goodman a été présenté par Nelson Goodman dans un article paru en 1946 dans la revue *Journal of Philosophy*. Goodman y expose son paradoxe de la manière suivante[1]. On considère une urne qui contient 100 boules. Chaque jour, une boule est extraite de l'urne durant 99 jours, jusqu'à aujourd'hui. A chaque tirage, il s'avère que la boule prélevée dans l'urne est rouge. A ce stade, on s'attend, de manière intuitive, à ce que la 100ème boule tirée soit également rouge. Cette prédiction est basée sur la généralisation selon laquelle toutes les boules présentes dans l'urne sont rouges. Le raisonnement sur lequel est basée cette dernière conclusion constitue une *induction énumérative*.

On peut traduire le raisonnement inductif précédent de manière plus formelle de la façon suivante. Soit R le prédicat *rouge*. Soient également $b_1$, $b_2$, $b_3$, ..., $b_{100}$ les 100 boules dans l'urne ($\wedge$ dénotant le connecteur logique *et*).

| | | |
|---|---|---|
| (1) | $Rb_1 \wedge Rb_2 \wedge Rb_3 \wedge ... \wedge Rb_{99}$ | *énumération* |

[1] Avec quelques adaptations mineures.

| | | |
|---|---|---|
| (2) | $Rb_1 \wedge Rb_2 \wedge Rb_3 \wedge ... \wedge Rb_{99} \wedge Rb_{100}$ | *de (1), induction* |
| (3) | $\therefore Rb_{100}$ | *de (2)* |

A ce stade, si on considère maintenant la propriété S « tiré avant aujourd'hui et rouge ou tiré après aujourd'hui et non-rouge » , on constate que cette propriété est également vérifiée par les 99 instances déjà observées. Mais la prédiction qui en résulte cette fois, basée sur la généralisation selon laquelle toutes les boules sont S, est que la 100ème boule sera non-rouge. Et ceci est contraire à la conclusion précédente, qui est elle-même pourtant conforme à notre intuition. Le raisonnement correspondant peut être ainsi détaillé :

| | | |
|---|---|---|
| (4) | $Sb_1 \wedge Sb_2 \wedge Sb_3 \wedge ... \wedge Sb_{99}$ | *énumération* |
| (5) | $Sb_1 \wedge Sb_2 \wedge Sb_3 \wedge ... \wedge Sb_{99} \wedge Sb_{100}$ | *de (4), induction* |
| (6) | $\therefore Sb_{100}$ | *de (5)* |

Mais ici, la conclusion selon laquelle la 100ème boule est S équivaut au fait que cette dernière sera non-rouge. Or ceci est en contradiction avec la conclusion résultant du raisonnement inductif précédent selon laquelle la 100ème boule sera rouge. Le paradoxe émerge ici à cause du fait que les deux conclusions (3) et (6) sont contradictoires. Intuitivement, l'application de l'énumération inductive à (4) paraît erronée. Mais la difficulté réside ici dans le fait de localiser avec précision où se trouve l'erreur de raisonnement à l'origine de cette fausse conclusion.

Goodman donne aussi dans son ouvrage *Faits, fictions et prédictions*, paru dans sa version originale en 1954, une version légèrement différente de son paradoxe, appliquée cette fois aux émeraudes :

> Supposez que toutes les émeraudes examinées avant un certain temps *t* aient été vertes. Dans ce cas, au

> temps *t*, nos observations confirment l'hypothèse selon laquelle toutes les émeraudes sont vertes ; et ceci est en accord avec notre définition de la confirmation [... ] Maintenant laissez-moi introduire un autre prédicat moins familier que « vert ». C'est le prédicat « vleu » et il s'applique à toutes les choses examinées avant *t* si elles sont vertes mais aux autres choses si elles sont bleues. Ainsi au temps *t* nous avons, pour chaque constatation matérielle rapportant qu'une émeraude donnée est verte, une constatation matérielle rapportant de manière parallèle que l'émeraude est « vleu ».

Cette version du paradoxe de Goodman est célèbre et basée sur le prédicat « vleu » [2]. La définition de « vleu » est la suivante : *vert et observé avant T ou non-vert et observé après T*. Il en résulte deux types de raisonnements concurrents. Un premier raisonnement met en œuvre une énumération inductive classique : à partir de l'observation selon laquelle toutes les émeraudes observées avant T étaient *vertes*, on conclut que la prochaine émeraude observée sera également verte (V dénotant *vert*, et $e_1$, $e_2$, $e_3$, ..., $e_{100}$ dénotant les émeraudes) :

| | | |
|---|---|---|
| (7) | $Ve_1 \wedge Ve_2 \wedge Ve_3 \wedge ... \wedge Ve_{99}$ | |
| (8) | $Ve_1 \wedge Ve_2 \wedge Ve_3 \wedge ... \wedge Ve_{99} \wedge Ve_{100}$ | *énumération de (7),* |
| (9) | $\therefore Ve_{100}$ | *induction de (8)* |

Le raisonnement alternatif est basé sur le même type d'énumération inductive appliqué au prédicat « vleu ». Du fait que toutes les émeraudes observées avant T étaient « vleues » , on conclut cette fois que la prochaine émeraude observée sera également « vleue » (« vleu » étant dénoté par G) :

[2] Dans le texte original: *grue*.

| | | |
|---|---|---|
| (10) | $Ge_1 \wedge Ge_2 \wedge Ge_3 \wedge ... \wedge Ge_{99}$ | |
| (11) | $Ge_1 \wedge Ge_2 \wedge Ge_3 \wedge ... \wedge Ge_{99} \wedge Ge_{100}$ | *énumération de (10),* |
| (12) | $\therefore Ge_{100}$ | *induction de (11)* |

Il s'ensuit alors une contradiction, puisqu'en vertu de (9) la 100ème émeraude sera verte, alors qu'il résulte de (11) que la 100ème émeraude sera non-verte. Les deux problèmes présentés par Goodman constituent deux variations du même paradoxe, car le prédicat S utilisé par Goodman dans son article de 1946 présente avec « vleu », une *structure* commune. En effet, P et Q étant deux prédicats, cette dernière structure correspond à la définition : (P et Q) ou (non-P et non-Q).

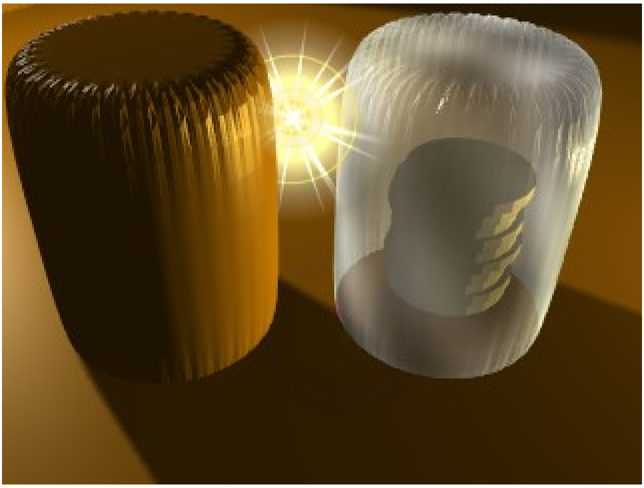
Le paradoxe de Goodman a engendré une énorme littérature et de nombreuses solutions de nature différente ont été proposées pour le résoudre. Goodman a ainsi proposé lui-même une telle solution, qui est basée sur la notion d'*enfouissement*[3]. Goodman, dans *Faits, fictions et prédictions* considère ainsi que le problème se ramène à celui d'établir une distinction entre les prédicats qui sont *projetables*, et ceux qui ne le sont pas. Les prédicats projetables peuvent valablement servir de support à une induction énumérative, alors que les autres, au nombre desquels se trouve « vleu », ne conviennent pas pour cela. Selon Goodman, les prédicats projetables sont ceux qui sont intégrés, enfouis dans notre pratique inductive courante. Il s'agit là d'un usage inductif qui se trouve ainsi avalisé par la pratique. Les prédicats projetables sont ceux qui sont en quelque sorte validés par l'usage courant, commun et passé. A l'inverse, les prédicats non projetables tels que « vleu » ne sont pas adaptés à l'usage inductif. Cependant, la solution de Goodman basée sur l'enfouissement dans le langage et l'usage courant ne s'est pas révélée satisfaisante. Car il s'avère que de nouveaux prédicats apparaissent chaque jour. De nombreux

3 *Entrenchment*.

néologismes sont en effet créés, qui s'intègrent très vite dans le langage courant et dans la pratique inductive commune. Même le prédicat « vleu » à l'origine si décrié nous est devenu quelque peu familier.

Une autre solution qui a notamment été proposée pour résoudre le paradoxe de Goodman est basée sur le fait que le prédicat « vleu » comporte une référence *temporelle*, à la différence du prédicat « vert ». Selon ce type de solution, il convient de ne pas utiliser pour l'induction des prédicats tels que « vleu », qui comportent de telles clauses temporelles. Toutefois, ce type de solution s'est avéré trop restrictif, car il existe des prédicats qui comportent une référence temporelle mais dont la projection inductive ne pose aucun problème. Considérons ainsi une tomate : celle-ci elle est verte avant maturité, et rouge après. Une telle propriété s'applique aux 99 tomates que je viens de trouver dans mon jardin, mais aussi à la 100ème tomate qui se trouve dans le jardin de mon voisin. En second lieu, il s'avère tout à fait possible de construire une version du paradoxe de Goodman qui est dépourvue d'une telle clause temporelle. Il suffit alors de construire un prédicat G basé par exemple sur l'association couleur-espace, en remplacement de l'association couleur-temps, pour obtenir une variation du paradoxe de Goodman qui s'affranchit d'une référence temporelle. Enfin, la réponse apportée par Nelson Goodman lui-même par rapport à ce type d'objection est que le prédicat « vert » peut également être défini avec une référence temporelle si l'on utilise « vleu » comme concept primitif. Il suffit ainsi de mettre en parallèle d'une part les prédicats « vert » et « bleu » et d'autre part « vleu » (vert avant T et bleu après T) et « bert » (bleu avant T et vert après T). Dans ce cas, il est tout à fait possible de définir « vert » et « bleu » à l'aide des notions primitives de « vleu » et « bert ». Un objet « vert » est alors défini comme « vleu » avant T et « bert » après T ; et de même, un objet « bleu » est défini comme « bert » avant T et « vleu » après T. Ainsi, les définitions de « vert », « bleu » et d'autre part « vleu », « bert » se révèlent parfaitement symétriques et comportent de manière identique une référence temporelle.

## 6. Le problème de Newcomb



Le problème de Newcomb a été décrit en 1960 par le physicien William Newcomb et a été introduit ensuite dans la littérature philosophique à travers un essai publié en 1969 par Robert Nozick. On peut décrire le problème de Newcomb de la manière suivante. Deux boites, A et B, se trouvent placées devant vous. L'une d'entre elles – la boite A – est transparente et contient 1000 euros. Vous êtes placé devant le choix suivant : soit prendre uniquement le contenu de la boite B ; soit prendre à la fois le contenu de la boite A et de la boite B. Vous savez également qu'un devin, dont les prédictions se sont révélées extrêmement fiables jusqu'à présent, placera un million d'euros dans la boite B s'il prédit que vous ne prendrez que cette dernière. En revanche, s'il prédit que vous prendrez à la fois les boites A et B, le devin laissera la boite B vide. Maintenant, choisissez-vous de prendre uniquement la boite B, ou bien de prendre les boites A et B ? En vertu d'un premier raisonnement (I), il apparaît que les prédictions effectuées dans le passé par le devin se sont révélées très fiables, et il n'y a pas de raison pour que la prédiction qu'il va effectuer avec vous ne se vérifie pas une fois de plus. Par conséquent, il apparaît prudent de ne prendre que la boite B, de manière à encaisser un million d'euros, ce qui représente déjà une très belle somme. A ce stade, il apparaît cependant qu'un raisonnement alternatif (II) peut également être tenu. Car au moment où vous préparez à ouvrir la boite B ou les deux boites, le devin a déjà effectué son choix. Par conséquent, si le devin a prédit que vous ouvrirez uniquement la boite B, il a alors placé un million d'euros dans la boite A. Ne serait-il alors pas absurde de laisser les 1000 euros qui se trouvent dans la boite A. Car cette dernière boite est transparente, et vous pouvez en observer le contenu. Vous raisonnez, et vous constatez que cela ne peut plus affecter le choix du devin. Par conséquent, mieux vaut ouvrir les deux boites, et encaisser ainsi 1001000 euros. A ce stade, il apparaît que chacun des deux raisonnements (I) et (II) semble fondé. Pourtant, tous deux conduisent à des conclusions contradictoires. Et l'énigme posée par le problème de Newcomb est précisément de savoir lequel des raisonnements (I) et (II) est valable.

Il est intéressant de formaliser quelque peu les données du problème de Newcomb, de manière à mettre en évidence certains éléments de sa structure interne. Il apparaît ainsi que la structure de l'énoncé est celle d'un double conditionnel :

(1) si <le devin prédit que le sujet ouvrira la boite B> alors <le devin place 1000000 euros dans la boite B>

(2) si <le devin prédit que le sujet ouvrira les boites A et B> alors <le devin place 0 euro dans la boite B>

De même, le raisonnement (I) peut être décrit de manière détaillée de la façon suivante :

(3) les prédictions effectuées dans le | *prémisse*

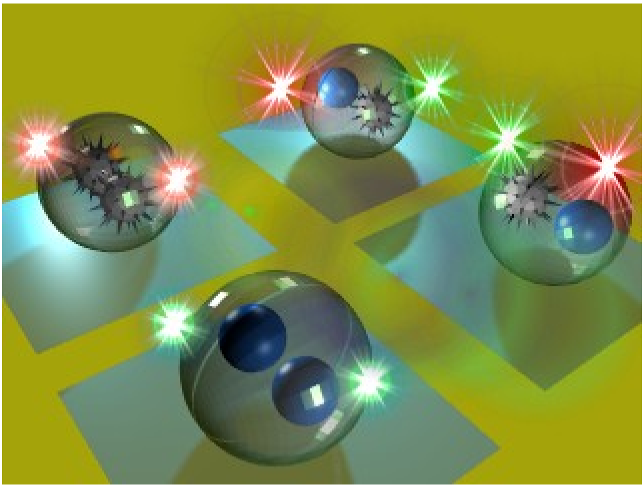
| | | |
|---|---|---|
| | passé par le devin se sont révélées très fiables | |
| (4) | les prédictions effectuées par le devin sont très fiables | *généralisation* |
| (5) | cette fois également, le devin devrait prédire mon choix | *de (4), induction* |
| (6) | si le devin a prédit que j'ouvrirai uniquement la boite B, alors il a placé 1000000 euros dans la boite B | *de (1)* |
| (7) | si le devin a prédit que j'ouvrirai les boites A et B alors il a placé 0 euro dans la boite B | *de (2)* |
| (8) | ∴ j'ai intérêt à ouvrir la boite B | *de (6),(7)* |

Et on peut de même formaliser ainsi le raisonnement (II) :

| | | |
|---|---|---|
| (9) | au moment ou j'effectue mon choix, les sommes d'argent sont déjà placées dans les boites, et celles-ci ne seront pas affectées par mon choix | *prémisse* |
| (10) | si le devin a placé 1000000 euros dans la boite A, alors en prenant également la boite B, je gagnerai 1001000 euros au lieu de 1000000 euros | *de (9)* |
| (11) | si le devin a placé 0 euro dans la boite A, alors en prenant également la boite B, je gagnerai 1000 euros | *de (9)* |
| (12) | dans les deux cas, j'obtiens un gain supérieur en prenant également la boite A | *de (10), (11)* |
| (13) | ∴ j'ai intérêt à ouvrir les boites A et B | *de (12)* |

Le paradoxe de Newcomb a donné lieu à un formidable engouement et a engendré une vaste littérature. Parmi les solutions qui ont été proposées pour résoudre le paradoxe, l'une d'elles met l'accent sur le fait que la situation correspondant au paradoxe est en réalité impossible et s'avère telle qu'on ne peut la rencontrer en pratique. Selon cette analyse, la partie de l'énoncé selon laquelle le devin peut prédire avec précision le choix du sujet n'est pas vraisemblable. En vertu de cette analyse, une telle clause fait appel à des propriétés extravagantes qui ne sont pas celles de notre monde physique, telles que la causalité rétrograde (la fait qu'un effet puisse agir sur sa propre cause) ou l'absence de libre-arbitre des individus. Une telle solution, cependant, ne s'est pas avérée satisfaisante. Car s'il est permis de mettre en doute l'existence de la causalité rétrograde ou l'absence de libre-arbitre, on peut néanmoins mettre en évidence d'autres variations du paradoxe de Newcomb qui ne font pas appel à de telles propriétés singulières. Il suffit pour cela de considérer une version probabiliste du paradoxe où la prédiction du devin est le plus souvent exacte. Car le devin pourrait bien se fonder sur des considérations d'ordre purement psychologique. Une étude menée sur le paradoxe de Newcomb a montré que 70% des gens choisissent de ne prendre que la boite B. Le devin pourrait ainsi posséder d'un programme d'ordinateur simulant de manière très performante le comportement et la psychologie humaine face à ce type de situation et effectuer ses prévisions en conséquence. Dans ce contexte, la clause de l'énoncé selon laquelle les prévisions du devin sont très souvent exactes serait tout à fait respectée.

## 7. Le dilemme du prisonnier



Le *dilemme du prisonnier* a été décrit par Merrill Flood et Melvin Dresher en 1950. Il peut être formulé de la manière suivante. Deux prisonniers, Jean et Pierre, sont interrogés par un juge qui les soupçonne d'avoir commis un crime. Le juge propose à chacun d'eux le marché suivant : « Vous disposez de deux possibilités : soit avouer, soit ne pas avouer. Mais attention, le choix que vous effectuerez aura une conséquence très importante sur la peine qui vous sera infligée. Ainsi, si l'un d'entre vous avoue mais que l'autre n'avoue pas, celui qui aura avoué sera libre alors que celui qui aura refusé d'avouer se verra infliger 10 ans de prison. En revanche, si vous avouez tous les deux, chacun d'entre vous n'aura que 5 ans de prison. Enfin, si aucun de vous n'avoue, je vous infligerai à tous les deux 1 an de prison. Maintenant, réfléchissez, puis déterminez-vous. Je vous ferai ensuite connaître ma sentence ».

A ce stade, il apparaît utile de décrire plus en détail la structure du dilemme du prisonnier. Il s'avère ainsi que les quatre cas suivants sont possibles :

(a) Jean avoue et Pierre avoue
(b) Jean avoue et Pierre n'avoue pas
(c) Jean n'avoue pas et Pierre avoue
(d) Jean n'avoue pas et Pierre n'avoue pas

De plus, l'annonce du juge peut être décrite à l'aide de la matrice suivante, qui définit les peines attribuées à chacun des deux prisonniers en fonction de leur attitude :

| | | |
|---|---|---|
| (a) | Jean avoue et Pierre avoue | *Jean : 5 ans*<br>*Pierre : 5 ans* |
| (b) | Jean avoue et Pierre n'avoue pas | *Jean : 0 an*<br>*Pierre : 10 ans* |
| (c) | Jean n'avoue pas et Pierre avoue | *Jean : 10 ans*<br>*Pierre : 0 an* |
| (d) | Jean n'avoue pas et Pierre n'avoue pas | *Jean : 1 an*<br>*Pierre : 1 an* |

Le problème inhérent au dilemme du prisonnier provient du fait que deux types de raisonnements différents apparaissent tous deux valables. En effet, en vertu d'un premier type (I) de raisonnement, il apparaît que le fait de ne pas avouer est ce qui donne à chacun le maximum de chances d'être libre. En effet, si l'un des prisonniers avoue, il en résulte une peine qui est de 5 ans (si l'autre avoue également) ou nulle (si l'autre n'avoue pas) ; ainsi, la peine qui en résulte est en moyenne de 2,5 ans : (5 + 0) / 2. En revanche, si le prisonnier n'avoue pas, il s'ensuit une peine de 10 ans (si l'autre avoue) ou de 1 an (si l'autre n'avoue pas également) ; ainsi, il en résulte une peine qui est en moyenne de 5,5 ans : (10 + 1) / 2. Il apparaît donc beaucoup plus rationnel d'avouer.

Cependant, un autre type de raisonnement apparaît également possible. Selon un autre point de vue (II) en effet, il s'avère que le fait de ne pas avouer se révèle très intéressant pour chacun des deux prisonniers. Car il n'en résulte qu'une peine d'un an pour chacun d'eux. Finalement, on se trouve en présence d'un dilemme, car chacune des options qui résulte des deux raisonnements (I) et (II) en compétition se révèle, d'un certain point de vue, optimale.

Le dilemme du prisonnier correspond à une situation concrète, pratique, qui possède des répercussions dans le domaine de la théorie des jeux, de l'économie, de la science politique, de la biologie, etc. Au niveau de la théorie des jeux, on distingue ainsi classiquement entre les jeux à somme nulle et ceux à somme non nulle. Pour les jeux à somme nulle, il existe un gagnant et un perdant, mais pas de situation intermédiaire (tel est le cas par exemple pour le tennis). A l'inverse, pour les jeux à somme non nulle, il existe un gagnant, un perdant, et une ou plusieurs situations intermédiaires (les échecs, où la possibilité de la partie nulle existe, en constituent un exemple). Dans ce contexte, le dilemme du prisonnier apparaît comme un jeu à somme non nulle, puisqu'il existe deux cas où les deux prisonniers reçoivent une peine identique : (1) s'ils avouent tous les deux ; et (2) s'ils n'avouent pas tous les deux.

On peut observer que le dilemme du prisonnier donne lieu à une importante variation lorsque le dilemme est répété. Il s'agit alors du *dilemme itéré du prisonnier*. Dans ce contexte, plusieurs stratégies apparaissent alors possibles. Il en résulte ainsi les stratégies élémentaires suivantes : *toujours avouer*, ou bien *ne jamais avouer*. Mais d'autres stratégies plus complexes sont possibles, basées notamment sur l'option choisie par l'autre prisonnier lors des coups précédents. Dans ce cas, les itérations conduisent alors à analyser la succession de coups joués par le prisonnier comme un type de comportement. A ce stade, les possibilités deviennent multiples. Une stratégie qui s'est avérée très performante a ainsi été dénommée *tit-for-tat*. La stratégie sur laquelle elle est basée est la suivante : avouer au premier coup, puis jouer au coup $n + 1$ ce qu'a joué l'autre prisonnier au coup $n$. Pour le dilemme itéré du prisonnier, il n'existe pas non plus de stratégie dont on puisse dire, de manière certaine, qu'elle est meilleure que les autres.

## 8. Le paradoxe de Cantor

Le paradoxe de Cantor a été découvert par Georg Cantor en 1899, mais n'a toutefois été publié qu'en 1932. L'idée générale du paradoxe réside dans le fait que la prise en considération de *l'ensemble de tous les ensembles* conduit à une contradiction. En effet, si l'on appelle C *l'ensemble de tous les ensembles*, il s'ensuit alors qu'il existe un ensemble C*, qui est lui-même défini comme l'ensemble composé des parties de l'ensemble C. Par définition, l'ensemble C qui est *l'ensemble de tous les ensembles* inclut donc également l'ensemble C*. Ceci implique que le cardinal – c'est-à-dire le nombre d'éléments – de l'ensemble C est supérieur ou égal au cardinal de l'ensemble C*. Or un théorème, établi par Cantor, établit qu'étant donné un ensemble E, le cardinal de E est inférieur au cardinal de l'ensemble E*, qui est constitué de toutes les parties de E. Ainsi, en vertu du théorème de Cantor, il s'ensuit que le cardinal de l'ensemble C*, qui inclut toutes les parties de C, est nécessairement plus grand que le cardinal de l'ensemble C. Il en résulte donc une contradiction.

Le raisonnement correspondant au paradoxe de Cantor peut être ainsi détaillé de manière plus formelle (*card* dénote ici le cardinal d'un ensemble) :

| | | |
|---|---|---|
| (1) | C est l'ensemble de tous les ensembles | *définition* |
| (2) | C* est l'ensemble de toutes les parties de l'ensemble C | *prémisse* |
| (3) | card (C) ≥ card (C*) | *de (1)* |
| (4) | pour tout ensemble E, l'ensemble E* de toutes les parties de E est tel que card (E) < card (E*) | *théorème de Cantor* |
| (5) | pour l'ensemble C, l'ensemble C* de toutes les parties de C est tel que card (C) < card (C*) | *de (4)* |
| (6) | ∴ card (C) ≥ card (C*) et card (C) < card (C*) | *de (3),(5)* |

Le paradoxe de Cantor appartient, de même que le paradoxe de Russell, à la catégorie des paradoxes ensemblistes. A l'instar du paradoxe de Russell, il apparaît au sein de la théorie naïve des ensembles, où la construction de l'ensemble C de tous les ensembles se trouve autorisée. Dans la théorie actuelle des ensembles, celle de Zermelo-Fraenkel, le paradoxe est évité car on ne peut construire l'ensemble C. En effet, un des axiomes de la théorie de Zermelo-Fraenkel, l'axiome de compréhension, a été conçu de manière plus restrictive que dans la théorie naïve des ensembles, afin d'interdire la construction de l'ensemble C de tous les ensembles. Mais une telle démarche peut paraître *ad hoc*, c'est-à-dire qu'il s'agit d'une restriction de la théorie des ensembles qui a pour seul but d'éviter les paradoxes et la contradiction qui en résulte. Dans ce contexte, de même que pour le

paradoxe de Russell, on ne peut considérer véritablement que l'on dispose actuellement d'une solution authentique pour le paradoxe de Cantor.

# 9. Le paradoxe de Grelling


Ce paradoxe a été inventé par Kurt Grelling. Il est également appelé *paradoxe des mots hétérologiques*. Le paradoxe de Grelling peut être énoncé de la manière suivante : certains adjectifs décrivent des propriétés qui s'appliquent à eux-mêmes, tels que « polysyllabique », « français ». De tels adjectifs peuvent être qualifiés d'*autologiques*. D'autres adjectifs, à l'inverse, décrivent des propriétés qui ne s'appliquent pas à eux-mêmes. Par exemple, « long », « italien », « monosyllabique ». On peut qualifier de tels mots d'*hétérologiques*. Ceci conduit à classer les mots en deux catégories : (a) autologiques ; (b) hétérologiques. Une telle distinction conduit toutefois à un paradoxe. Compte tenu des définitions précédentes, le paradoxe apparaît en effet lorsqu'on s'interroge sur le statut du prédicat *hétérologique* lui-même. Ainsi, « hétérologique » est-il autologique ou bien hétérologique ? Car si « hétérologique » est hétérologique, alors par définition, « hétérologique » est autologique. Et inversement, si « hétérologique » est autologique, il en résulte qu'il est hétérologique. La conclusion est paradoxale, car il s'ensuit qu'« hétérologique » est hétérologique si et seulement s'il est autologique.

Les définitions et le raisonnement qui conduisent au paradoxe de Grelling peuvent être présentées de manière plus détaillée de la manière suivante (H et ~H dénotant respectivement *hétérologique* et non-hétérologique – c'est-à-dire *autologique* – et ϕ dénotant une propriété donnée) :

| | | |
|---|---|---|
| (1) | H(« ϕ ») si et seulement si ~ϕ(« ϕ ») | *définition 1* |
| (2) | ~H(« ϕ ») si et seulement si ϕ(« ϕ ») | *définition 2* |
| (3) | si H(« H ») | *hypothèse 1* |
| (4) | alors ~H(« H ») | *de (1)* |
| (5) | si ~H(« H ») | *hypothèse 2* |
| (6) | alors H(« H ») | *de (2)* |
| (7) | ∴ H(« H ») si et seulement si ~H(« H ») | *de (3),(4), (5),(6)* |

Et il apparaît que l'on ne peut attribuer valablement au prédicat « hétérologique » ni la propriété *hétérologique* ni la propriété *autologique*.

A ce stade, il est intéressant d'étudier également le statut du mot « autologique » lui-même. Ainsi, « autologique » est-il hétérologique ou bien autologique ? Le raisonnement concernant « autologique » s'établit comme suit :

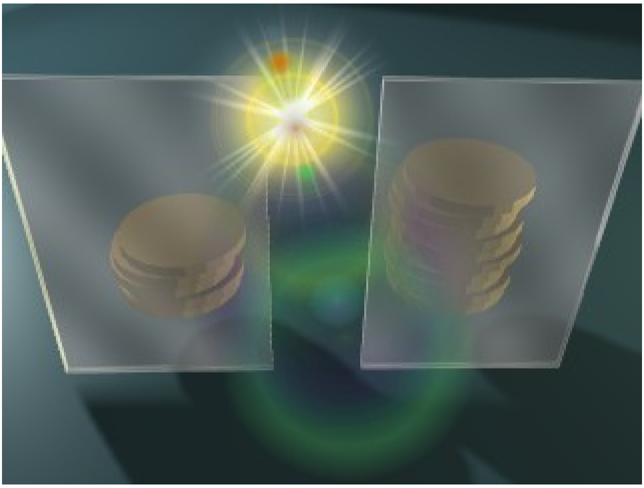
| | | |
|---|---|---|
| (1) | H(« ϕ ») si et seulement si ~ϕ(« ϕ ») | *définition 1* |
| (2) | ~H(« ϕ ») si et seulement si ϕ(« ϕ ») | *définition 2* |
| (8) | si H(« ~H ») | *hypothèse 1* |
| (9) | alors ~~H(« ~H ») | *de (1)* |
| (10) | alors H(« ~H ») | *de (9)* |
| (11) | si ~H(« ~H ») | *hypothèse 2* |
| (12) | alors ~H (« ~H ») | *de (2)* |

Ici, l'étape particulière (10) est justifiée par l'élimination de la double négation. Et dans ce cas, il apparaît que si « autologique » est hétérologique alors il est hétérologique ; et de même, si « autologique » est autologique alors il est autologique. Ainsi, il s'avère que l'on ne parvient pas non plus à déterminer valablement si « autologique » est hétérologique ou non.

Parmi les solutions qui ont été proposées pour résoudre le paradoxe de Grelling, l'une d'entre elles conduit à observer que la structure du paradoxe est très similaire à celle du paradoxe de Russell. Ainsi, les deux paradoxes présenteraient une structure commune et conduiraient à une solution de même nature.

Une autre solution conduit, de même que pour le paradoxe du Menteur, à rejeter les définitions de tous les prédicats qui présentent une structure auto-référentielle. Pourtant, une telle solution ne s'avère pas non plus satisfaisante. En effet, elle apparaît beaucoup trop restrictive, car il s'avère que l'on parvient tout à fait valablement à déterminer le statut de nombreux prédicats auto-référentiels tels que par exemple *polysyllabique*. Proscrire purement et simplement tous les prédicats dont la structure est auto-référentielle serait payer un prix beaucoup trop fort pour la seule élimination du paradoxe.

# 10. Le paradoxe des deux enveloppes



Le paradoxe des deux enveloppes s'énonce de la façon suivante : devant vous se trouvent deux enveloppes qui contiennent chacune une somme d'argent et vous savez de manière certaine que l'une d'entre elles contient le double de l'autre. Vous prenez l'une des deux enveloppes au hasard. Maintenant, vous avez le choix entre garder l'enveloppe que vous avez en main, ou bien échanger avec l'autre enveloppe. Que décidez-vous de faire ? Un premier type de raisonnement (I) vous vient immédiatement à l'esprit : la situation concernant chacune des deux enveloppes est tout à fait identique. En choisissant seulement l'une des deux enveloppes, vous n'avez obtenu aucune information nouvelle. Par conséquent, le choix de l'une ou l'autre est équivalent. Vous décidez donc de conserver l'enveloppe que vous avez initialement prise. Cependant, il apparaît qu'un autre type de raisonnement (II) se révèle également possible : soit $x$ la somme contenue dans l'enveloppe que vous avez entre les mains.

L'autre enveloppe contient donc une somme qui est égale soit à $2x$, soit à $1/2x$. Ces deux situations sont équiprobables et chacune d'elles peut se voir attribuer une probabilité de 1/2. Par conséquent, la probabilité générale peut être calculée ainsi : $2x$ x $1/2 + 1/2x$ x $1/2 = 5/4x$. Il s'ensuit que dans le cas général, l'autre enveloppe contient une somme égale à $5/4x$ c'est-à-dire 1,25 $x$. Ainsi, il s'avère que l'autre enveloppe contient une somme qui est d'un quart supérieure à celle que vous avez dans les mains. Par conséquent, vous avez intérêt à échanger avec l'autre enveloppe. Cependant, une fois l'enveloppe échangée, un raisonnement de même nature vous conduit à échanger à nouveau l'enveloppe, et ainsi de suite *ad infinitum*.

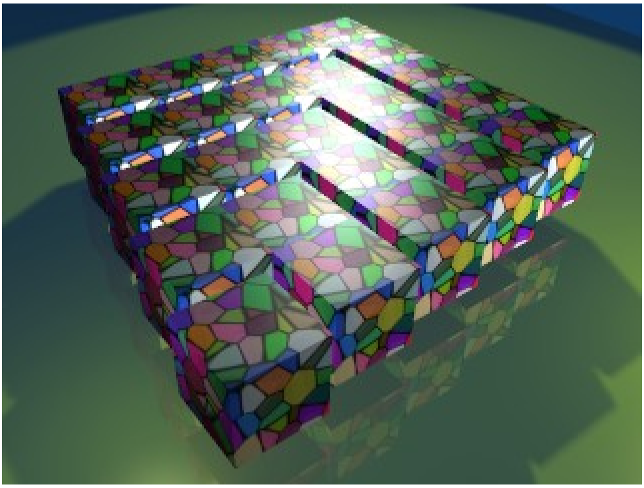
Dans le paradoxe des deux enveloppes, c'est clairement le raisonnement (II) qui est en cause, puisqu'il conduit à la conclusion absurde qu'il convient d'échanger les enveloppes à l'infini. Pourtant, la tâche qui consiste à déterminer avec précision l'étape fallacieuse dans le raisonnement (II) s'avère très difficile. A cette fin, il est utile de formaliser davantage les différentes étapes inhérentes au raisonnement (II) :

| | | |
|---|---|---|
| (1) | l'autre enveloppe contient soit (a) la somme $2x$ soit (b) la somme $1/2x$ | *prémisse* |
| (2) | la probabilité de chacune des situations (a) et (b) est 1/2 | *prémisse* |
| (3) | la probabilité générale est que l'autre enveloppe contienne: $2x$ x $1/2 + 1/2x$ x 1/2 | *de (1),(2)* |
| (4) | la probabilité générale est que l'autre enveloppe contient $1{,}25x$ | *de (3)* |
| (5) | ∴ j'ai intérêt à échanger avec l'autre enveloppe | *de (4)* |

Parmi les solutions qui ont été proposées pour résoudre le paradoxe, l'une d'elles fait valoir que l'assertion (2) selon laquelle la seconde enveloppe contient $2x$ ou $1/2x$

avec une probabilité égale à 1/2, n'est pas vraie dans tous les cas. Ainsi, Franck Jackson et ses coauteurs ont fait valoir dans un article publié en 1994 qu'en réalité, les valeurs de *x* et les paires de valeurs qui en résultent n'ont pas toutes la même probabilité de se trouver dans les enveloppes. En effet, il existe certaines valeurs limites – soient très petites, soit très grandes – que l'on n'a que très peu de chances, pour des raisons pratiques, de rencontrer. Ainsi, les deux valeurs qui peuvent se trouver dans l'autre enveloppe ne sont pas équiprobables et par conséquent, la prémisse (2) n'est pas exacte. Toutefois, une telle solution n'est pas apparue satisfaisante. En effet, ainsi que l'on fait remarquer McGrew et ses coauteurs dans un article paru en 1997, on parvient à faire resurgir le paradoxe en considérant une variante de ce dernier, où dans les enveloppes ne sont pas placées des sommes d'argent, mais de simples morceaux papier où sont inscrits des nombres.

# 11. Le paradoxe de Moore



Le *paradoxe de Moore* a été décrit par G. E. Moore dans un texte paru en 1942. Si l'on considère ainsi la proposition suivante :

(1) Il pleut et je ne crois pas qu'il pleut

il s'ensuit qu'une telle proposition est a priori absurde. Intuitivement, une telle proposition présente une nature contradictoire. Pourtant, il s'avère qu'il existe certaines situations où une assertion telle que (1) peut être valablement exprimée. Une telle situation correspond par exemple au cas où une personne possède une croyance justifiée qu'un événement donné ne surviendra pas, mais où cet événement survient finalement, en rendant finalement fausse la croyance initiale. Ainsi, une personne peut croire fermement qu'il ne pleut pas aujourd'hui en se basant sur des prévisions météo entendues la veille, alors qu'il pleut en réalité. Dans ce contexte, l'assertion (1) apparaît alors à nouveau plausible.

Il s'avère utile ici d'analyser plus en détail la structure de (1). Si l'on considère ainsi une proposition quelconque P, il s'ensuit que l'assertion (1) présente la structure suivante :

(2) P et je ne crois pas que P

On le voit, la structure logique de (2) est la suivante (Q dénotant « je crois » et ~ la négation) :

(3) $P \wedge \sim Q(P)$

On distingue habituellement deux variations du paradoxe de Moore : le paradoxe de Moore de Hintikka, et le paradoxe de Moore de Wittgenstein. Le paradoxe de Moore de Hintikka présente une structure qui est celle de (2) et correspond à la version originale du paradoxe de Moore. En revanche, le paradoxe de Moore de Wittgenstein porte sur la proposition :

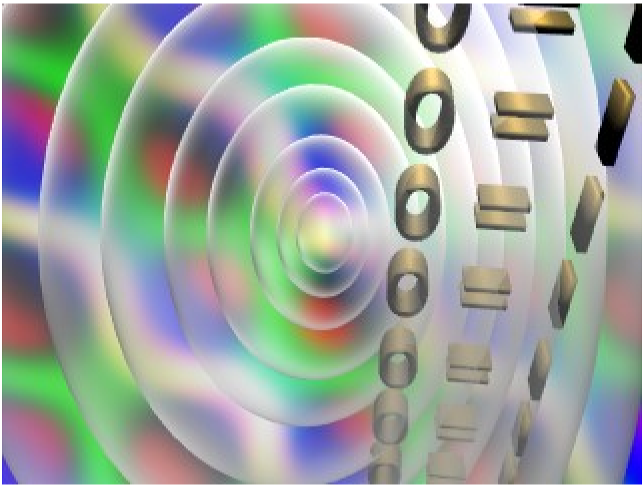
(4) P et je crois que non-P

qui présente la structure logique :

(5) $P \wedge Q(\sim P)$

Selon certains auteurs, le paradoxe de l'examen-surprise s'assimile au paradoxe de Moore. Tel a été notamment le point de vue émis par Robert Binkley, dans un article publié en 1968, où il a fait valoir que si la période dans laquelle l'examen peut avoir lieu n'est que d'un jour, l'annonce du professeur présente alors la structure du paradoxe de Moore. Car l'annonce du professeur faite aux étudiants est alors la suivante : « Il y aura un examen demain mais vous ne saurez pas que cet examen aura lieu demain ». Dès lors que les étudiants concluent que l'examen ne peut avoir lieu, ils se trouvent alors, le jour-même de l'examen, dans une situation qui permet à l'annonce du professeur d'être validée. Et il en résulte alors une situation réelle qui correspond, sans contradiction, à la proposition (1).

# 12. Le paradoxe de Löb



Le paradoxe de Löb est mentionné dans l'ouvrage *The Liar*, de Jon Barwise et John Etchemendy, paru en 1987. Les auteurs indiquent que le paradoxe a été porté à leur attention par Dag Westerstahl. Le paradoxe de Löb, à partir d'une proposition qui semble inoffensive, conduit à la conclusion dévastatrice que toute proposition est vraie. La proposition qui constitue le point de départ du raisonnement est la suivante :

| | | |
|---|---|---|
| (1) | si la proposition (1) est vraie, alors 0 = 1 | *prémisse* |

Une telle proposition présente la structure d'une proposition conditionnelle (c'est-à-dire qui revêt la forme : si <*antécédent*> alors <*conséquent*>) dont l'antécédent est « la proposition (1) est vraie » et le conséquent est « 0 = 1 ». Le paradoxe apparaît dès lors que l'on considère l'hypothèse selon laquelle l'antécédent de (1), c'est-à-dire « la proposition (1) est vraie », est vraie. Si l'antécédent de

(1) est vrai, il s'ensuit alors que 0 = 1. Mais cette dernière proposition n'est autre que (1) elle-même. Il en résulte donc, par application du *modus ponens* (un principe logique en vertu duquel si P, P → Q, alors Q), que la proposition (1) elle-même est vraie. En conséquence, la proposition (1) vient d'être prouvée. Il s'agit là d'un cas d'application de *preuve conditionnelle*. Cependant, si (1) est vraie, une nouvelle application du *modus ponens* conduit enfin au fait que 0 = 1.

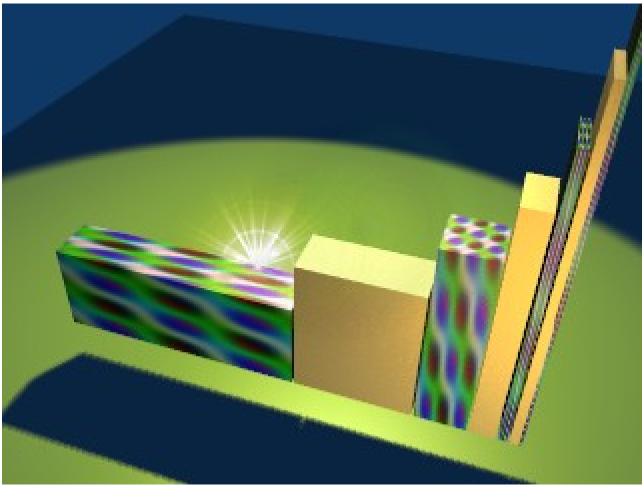
On peut décrire de manière plus détaillée les différentes étapes du raisonnement qui conduisent au paradoxe de Löb :

| | | |
|---|---|---|
| (1) | si la proposition (1) est vraie, alors 0 = 1 | *prémisse* |
| (2) | si la proposition (1) est vraie | *hypothèse* |
| (3) | alors 0 = 1 | *de (1),(2)* |
| (4) | si la proposition (1) est vraie, alors 0 = 1 | *de (2),(3)* |
| (5) | ∴ (1) est vraie | *de (4)* |
| (6) | ∴ 0 = 1 | *de (1),(5)* |

Le paradoxe de Löb conduit ainsi à prouver, à partir d'une proposition qui semble pourtant inoffensive, n'importe quelle proposition. De même que pour les autres paradoxes contemporains, la tâche qui consiste à déterminer la cause précise du paradoxe s'avère très difficile.

Une tentative de solution conduit à observer que la structure de (1) est auto-référentielle. Il s'agit là d'un point commun avec d'autres paradoxes, et en particulier le paradoxe du Menteur. Mais la solution qui consiste à interdire les propositions présentant une structure auto-référentielle ne convient pas non plus ici. En effet, il s'agit là d'une mesure trop radicale et restrictive, qui conduit à éliminer des propositions dont la structure est auto-référentielle, mais qui ne présentent pourtant pas de problème pour se voir attribuer une valeur de vérité. Ici encore se pose le problème de la définition du critère qui permet de distinguer entre : (a) les propositions auto-référentielles qui admettent valablement une valeur de vérité ; (b) les propositions auto-référentielles auxquelles on ne peut assigner valablement une valeur de vérité.

## 13. Le paradoxe de la course



Le *paradoxe de la course* constitue un des célèbres paradoxes dus à Zénon d'Elée. On en trouve la mention très claire dans la Physique d'Aristote :

> Tu ne peux pas franchir en un temps fini un nombre de points infini. Tu es obligé de franchir la moitié d'une distance donnée quelconque avant de franchir le tout, et la moitié de cette moitié avant de pouvoir franchir celle-ci. Et ainsi de suite ad infinitum, de sorte qu'il y a un nombre infini de points dans n'importe quel espace donné, et tu ne peux en toucher un nombre infini l'un après l'autre en un temps fini.

De manière informelle, le paradoxe peut être décrit de la façon suivante. Un coureur désire parcourir la distance qui sépare un point A d'un point B. Pour aller jusqu'à B, le coureur doit d'abord parcourir la moitié de la distance qui sépare le point A du point B. Mais une fois qu'il a parcouru la moitié de cette distance, le coureur doit encore parcourir la moitié de la distance qui le sépare de l'arrivée en B. Une fois arrivé à ce point, le coureur aura parcouru les trois-quarts de la distance qui le sépare de B. Mais de là, il devra encore parcourir la moitié de la distance le séparant de l'arrivée, et ainsi de suite *ad infinitum*. Ainsi, le coureur devra parcourir un nombre infini de fois des distances qui sont elles-mêmes finies. Or ceci devrait prendre un temps infini. Par conséquent, le coureur ne parviendra jamais en B. Il s'ensuit ainsi que tout mouvement est impossible.

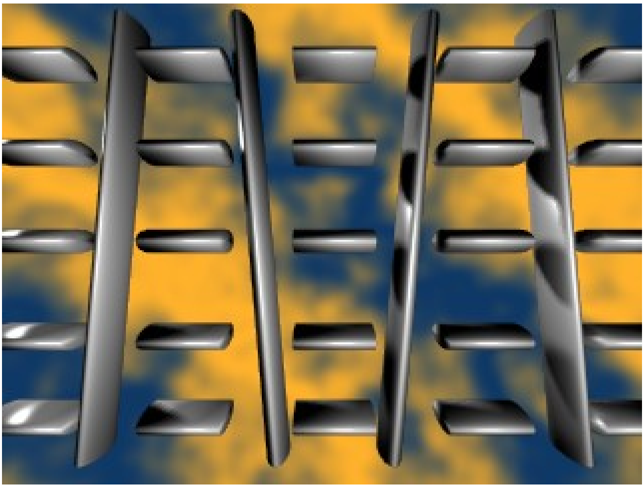
On peut décrire le paradoxe de manière un peu plus formelle. Soit *d* la distance séparant A de B. Dans ce cas, le coureur doit d'abord parcourir 1/2 de *d*, puis 1/4 de *d*, puis 1/8, puis 1/16, et ainsi de suite ad infinitum. Le raisonnement qui conduit au paradoxe de la course peut donc être décrit ainsi :

| | | |
|---|---|---|
| (1) | pour aller d'un point à un autre, un coureur doit d'abord parcourir la moitié de la distance qui sépare les deux points | *prémisse* |
| (2) | le coureur désire parcourir la distance *d* qui sépare le point A du point B | *prémisse* |
| (3) | pour aller de A à B, le coureur doit d'abord parcourir 1/2 *d* | *de (1),(2)* |
| (4) | une fois parvenu à 1/2 *d*, le coureur doit ensuite parcourir 1/4 *d* | *de (1),(2), (3)* |
| (5) | une fois parvenu à 3/4 *d*, le coureur doit ensuite parcourir 1/8 *d* | *de (1), (2),...,(4)* |
| (6) | ... | *de (1), (2),...,(5)* |
| (7) | le coureur devra parcourir un nombre infini de fois une fraction de *d* | *de (3), (4),...,(6)* |
| (8) | il est impossible de parcourir un nombre infini de distances en un temps fini | *prémisse* |
| (9) | ∴ le coureur ne parviendra jamais au point B | *de (7),(8)* |

Une premier type de réponse qui peut être apportée par rapport au paradoxe, est formulé par Aristote par l'intermédiaire de Simplicius : chacun sait par l'expérience individuelle que l'on peut se déplacer d'un point à un autre. Par conséquent, on peut également se déplacer d'un point A à un point B dans le cas correspondant à l'énoncé du paradoxe. Le coureur parviendra donc au point B, de la même manière que nous parvenons à l'endroit où nous souhaitons nous déplacer dans la vie courante. Une telle objection, toutefois, ne se révèle pas convaincante. En effet, la constatation empirique qu'elle met en évidence s'avère bien sûr vraie. Cependant, il s'agit précisément d'une des composantes du paradoxe. Car ce qui constitue ici le cœur du paradoxe, c'est que le raisonnement inhérent au paradoxe de la course conduit à une conclusion qui contredit les données courantes de l'expérience. Ainsi, cette objection ne fait que mettre l'accent sur un des éléments du paradoxe. Ce qui s'avère nécessaire en revanche, c'est de déterminer avec précision l'étape fallacieuse dans le raisonnement décrit par Zénon.

Une autre réponse, que beaucoup considèrent comme une résolution convaincante du paradoxe de la course, résulte directement des travaux de Cauchy et de sa théorie des séries infinies. En effet, Cauchy a montré que la somme d'une série infinie était parfois finie. En l'espèce, il s'avère que la somme de la série infinie $1/2 + 1/4 + 1/8 + 1/16 + ... + 1/2^n$ est égale à 1. Dans ces circonstances, chaque distance intermédiaire se trouve ainsi parcourue en un temps fini. La distance *d* est donc parcourue un temps fini, qui est égal à la somme des temps intermédiaires.

## 14. Le paradoxe de la pierre

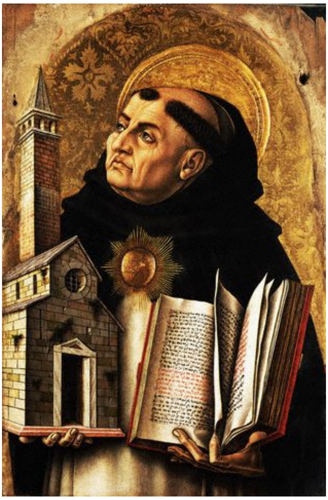


Le *paradoxe de la pierre* est un paradoxe qui trouve son origine dans les discussions sur la notion d'omnipotence initiées par Thomas d'Aquin. Dans sa forme moderne, le paradoxe de la pierre a été décrit par W. Savage en 1967, dans un article publié par la revue *Philosophical Review*. Il peut être énoncé de la manière suivante : soit tout d'abord la définition selon laquelle Dieu est un être omnipotent. Considérons ensuite une pierre qui présente la caractéristique suivante : elle est tellement lourde que Dieu ne peut la soulever. A ce stade, il existe deux possibilités : soit Dieu peut la créer, soit Dieu ne peut pas la créer. Envisageons tout d'abord la première hypothèse. Si Dieu peut créer une telle pierre, il s'ensuit donc que Dieu ne peut la soulever. Par conséquent, si Dieu peut créer une telle pierre, il existe ainsi une tâche que Dieu ne peut accomplir. Considérons maintenant la seconde hypothèse, en vertu de laquelle Dieu ne peut créer une telle pierre. Dans ce cas, il s'ensuit également qu'il existe une tâche que Dieu ne peut accomplir. Ainsi, la prise en compte de chacune des deux hypothèses conduit à la conclusion que dans chacun des cas, il existe une tâche que Dieu ne peut accomplir. Et ceci se révèle en contradiction avec le fait que Dieu est omnipotent. Il s'ensuit donc que Dieu n'existe pas.

Les étapes de l'argument peuvent être décomposées de la manière suivante :

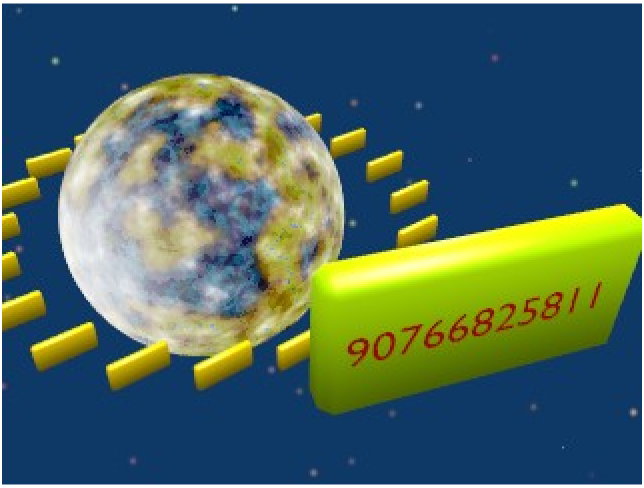
| | | |
|---|---|---|
| (1) | Dieu est un être omnipotent | *définition* |
| (2) | soit Dieu peut créer une pierre qu'il ne peut soulever, soit Dieu ne peut pas la créer | *dichotomie* |
| (3) | Dieu peut créer une pierre qu'il ne peut soulever | *hypothèse 1* |
| (4) | Dieu ne peut soulever une pierre | *de (3)* |
| (5) | il existe une tâche que Dieu ne peut accomplir | *de (4)* |
| (6) | Dieu ne peut pas créer une pierre qu'il ne peut soulever | *hypothèse 2* |
| (7) | il existe une tâche que Dieu ne peut accomplir | *de (6)* |
| (8) | il existe une tâche que Dieu ne peut accomplir | *de (5),(7)* |
| (9) | ∴ Dieu n'est pas un être omnipotent | *de (8)* |

Une solution qui a été formulée pour résoudre le paradoxe de la pierre repose sur le fait que la notion de *pierre que Dieu ne peut soulever* présente elle-même une nature contradictoire. Le statut d'une telle pierre, si elle existait, serait ainsi contradictoire par nature. Et il n'est donc pas étonnant que l'utilisation d'une notion contradictoire dans un argument entraîne des conséquences illogiques. La notion de *pierre que Dieu ne peut soulever* peut être ainsi comparée à un « cercle carré » ou à un « célibataire marié ». Car on peut en effet avoir exactement le même type d'argument avec un « cercle carré », conduisant de la même manière à une conséquence contradictoire.



Selon un autre point de vue, qui résulte des écrits de Thomas d'Aquin, le concept d'*omnipotence* ne peut pas être utilisé sans restriction. Car la notion d'omnipotence divine ne doit être envisagée que par rapport aux choses qui sont réellement possibles. En aucun cas, la notion d'omnipotence n'entraîne la capacité d'accomplir des choses impossibles. Un tel point de vue peut être appliqué directement au paradoxe de la pierre. Il s'ensuit alors que le fait de soulever une pierre que personne ne peut soulever, constitue précisément une tâche impossible.

## 15. L'argument de l'Apocalypse




L'argument de l'Apocalypse est un raisonnement qui a été énoncé par l'astrophysicien Brandon Carter, au début des années 1990. Ce type de raisonnement a également été découvert de manière indépendante par Richard Gott et H. Nielsen. L'argument de l'Apocalypse a ensuite été développé de manière détaillée et défendu par le philosophe canadien John Leslie dans une série de publications. La caractéristique principale de l'argument de l'Apocalypse et que les prémisses du raisonnement correspondant semblent tout à fait acceptables, alors que la conclusion se révèle inacceptable pour la plupart des gens.

Le raisonnement sur lequel est basé l'argument de l'Apocalypse est le suivant. On considère tout d'abord une urne qui comprend soit 10, soit 1000 boules. Les boules sont numérotées 1, 2, 3, 4, 5, .... Les hypothèses en compétition sont ainsi les suivantes :

(H1) l'urne comprend 10 boules numérotées
(H2) l'urne comprend 1000 boules numérotées

On considère que la probabilité initiale que l'urne contienne 10 boules ou 1000 boules est 1/2. Maintenant, vous tirez au hasard une boule dans l'urne et vous découvrez que celle-ci possède le numéro 5. Ce tirage rend-il plus probable l'hypothèse selon laquelle l'urne contient 10 boules, ou celle selon laquelle elle en contient 1000 ? Compte tenu de l'information nouvelle selon laquelle la boule extraite de l'urne porte le numéro 5, il apparaît qu'une révision à la hausse de la probabilité initiale de l'hypothèse selon laquelle l'urne contient seulement 10 boules, doit être effectuée. En effet, le tirage au hasard de la boule numéro 5 rend beaucoup plus probable cette dernière hypothèse. Car si l'urne ne contient que 10 boules, il existe 1 chance sur 10 de tirer la boule numéro 5. En revanche, si l'urne contient 1000 boules, il existe 1 chance sur 1000 de tirer la boule qui porte le numéro 5. Un calcul précis à l'aide du théorème de Bayes conduit à revoir à 0,99 la probabilité initiale que l'urne contienne 10 boules. Un tel raisonnement, basé sur le contenu d'une urne, se révèle consensuel.

A ce stade, on fait maintenant le parallèle avec la situation humaine. On considère ainsi deux hypothèses concernant l'évolution de l'humanité. On peut envisager ainsi que la population totale des humains ayant jamais existé atteindra soit 100 milliards, soit 10000 milliards. On formule ainsi les deux hypothèses suivantes concernant l'avenir de l'humanité :

(H3) l'humanité comptera au total 100 milliards d'humains
(H4) l'humanité comptera au total 10000 milliards d'humains

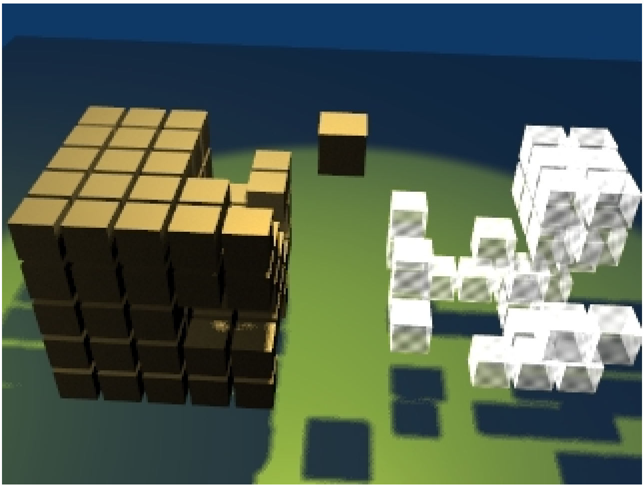
La première hypothèse correspond à une extinction prochaine et rapide de l'humanité, alors que la seconde correspond à une durée de vie très longue de l'humanité, qui pourrait ainsi coloniser d'autres planètes et s'étendre à travers la galaxie, etc. On attribue, pour simplifier, une probabilité de 1/2 à chacune de ces deux hypothèses. A ce stade, je suis amené à prendre en considération mon rang depuis la naissance de l'humanité. Considérant ainsi que je suis le 70000000000ème humain, je suis donc amené à raisonner de la même manière que je l'ai fait auparavant avec l'urne. Et ceci conduit à réviser à la hausse la probabilité initiale selon laquelle la population totale des humains ayant jamais existé n'atteindra que 100 milliards. Finalement, ceci plaide pour la probabilité – beaucoup plus grande qu'on ne l'aurait imaginé de prime abord – d'une extinction prochaine de l'humanité. Mais à la différence du cas précédent concernant l'urne, cette dernière conclusion apparaît cette fois tout à fait inacceptable et contraire à l'intuition. Dans le raisonnement qui a conduit à la conclusion selon laquelle l'humanité devrait rencontrer une extinction prochaine, une étape paraît être défectueuse. Mais la tâche de déterminer avec précision le point faible dans l'argument de l'Apocalypse s'avère une entreprise très difficile, pour laquelle les avis divergent considérablement.

Une première approche pour essayer de résoudre le problème posé par l'argument de l'Apocalypse est simplement d'accepter sa conclusion. Selon certains auteurs, et en particulier John Leslie, l'argument est correct et la conclusion qui en résulte doit être acceptée (avec une réserve importante toutefois, qui concerne le cas où notre univers n'est pas entièrement déterministe). Leslie se base pour cela sur le fait qu'il a réfuté, dans deux articles publiés en 1992 dans la revue *Mind* et dans son ouvrage *The End of the World* paru en 1996, de manière souvent convaincante, un nombre impressionnant d'objections à l'argument de l'Apocalypse. Cependant, l'acceptation de la conclusion de l'argument de l'Apocalypse demeure tout à fait contraire à l'intuition. D'autre part, l'acceptation que la simple connaissance de notre rang de naissance conduit à réévaluer à la hausse la probabilité de l'extinction prochaine de l'humanité, conduit à une conclusion de même nature dans nombre de situations courantes analogues. Il s'ensuit par exemple une révision à la hausse de la probabilité de la disparition prochaine de l'association à laquelle je viens d'adhérer, etc.

Un autre type de solution, que j'ai développée dans un article publié en 1999 par la revue *Canadian Journal of Philosophy*, consiste à considérer que la *classe de référence* sur laquelle porte l'argument de l'Apocalypse, c'est-à-dire l'espèce humaine, n'est pas définie avec précision. Car doit-on assimiler cette dernière à la sous-espèce *homo sapiens sapiens*, à l'espèce *homo sapiens*, au genre *homo*, etc. ? On peut ainsi choisir la classe de référence de manière différente, en opérant par restriction ou par extension. Dans l'énoncé de l'argument de l'Apocalypse, aucun critère objectif permettant de choisir la classe de référence, n'est présent. Il s'ensuit donc un choix arbitraire de cette dernière. Supposons alors que j'assimile, de manière arbitraire, la classe de référence à la sous-espèce homo sapiens sapiens. Il s'ensuit alors, par application de l'argument de l'Apocalypse, un décalage bayesien en faveur de l'hypothèse selon laquelle la sous-espèce homo sapiens sapiens est promise à une prochaine extinction. Toutefois, l'extinction de la sous-espèce homo sapiens sapiens peut aussi bien s'accompagner de l'apparition d'une ou plusieurs sous-espèces nouvelles, telles que *homo sapiens supersapiens*. Dans ce cas, la disparition de la classe de référence qui s'identifie, par restriction, à la sous-espèce homo sapiens sapiens, s'accompagne de la survie d'une classe de référence plus étendue, qui s'assimile à l'espèce homo sapiens. Un tel raisonnement a pour effet de rendre l'argument de l'Apocalypse inoffensif et d'en neutraliser la conclusion initialement dévastatrice. On peut objecter toutefois à une telle solution qu'elle admet toujours la validité de l'argument vis-à-vis d'une classe de référence restreinte telle qu'homo sapiens sapiens, alors même qu'une telle conclusion – bien qu'inoffensive – apparaît contraire à l'intuition.

Une autre solution qui a été proposée récemment par George Sowers, dans un article publié en 2002 dans la revue *Mind*, est la suivante. Selon l'auteur, l'analogie avec l'urne qui sous-tend l'argument de l'Apocalypse n'est pas valable, car notre rang de naissance individuel n'est pas obtenu de manière aléatoire comme le sont les numéros des boules extraites de l'urne. En effet, notre rang de naissance est indexé sur la position temporelle qui correspond à notre naissance. Par conséquent, conclut Sowers, le raisonnement qui sous-tend l'argument de l'Apocalypse est fallacieux, car il est basé sur une fausse analogie. Pourtant, l'analyse de Sowers n'est pas entièrement convaincante. En effet, on peut très bien imaginer une analogie avec une urne légèrement différente, où le tirage de la boule s'effectue de manière aléatoire, mais où le numéro de la boule est indexé sur la position temporelle correspondante. Il suffit pour cela de considérer un dispositif comportant une urne dont la boule n° $n$ par exemple se trouve extraite au hasard. Ensuite le mécanisme expulse la boule n° 1 au temps $T_1$, la boule n° 2 au temps $T_2$, la boule n° 3 au temps $T_3$, la boule n° 4 au temps $T_4$, ... et pour finir la boule n° $n$ au temps $T_n$. Le dispositif s'arrête alors. Et dans ce cas, il apparaît bien que le tirage de la boule a été effectué de manière aléatoire, alors même que le numéro de la boule est indexé sur la position temporelle correspondante.

## 16. Le problème du navire de Thésée



Dans la littérature, on trouve la trace pour la première fois du *problème du navire de Thésée* dans l'œuvre de Plutarque. Le problème peut être décrit de la manière suivante. Thésée possède un navire avec lequel il prend un jour la mer, accompagné de plusieurs de ses compagnons. Soit A ce dernier navire, qui est donc le « navire de Thésée ». Pendant le voyage, des avaries multiples rendent nécessaires de nombreuses réparations et c'est ainsi qu'assez souvent, des pièces du navire doivent être remplacées par des pièces neuves. De longues années s'écoulent ainsi et alors que l'heure du retour approche, il s'avère que toutes les pièces du navire ont finalement été remplacées. Ainsi, lors du retour de Thésée en Grèce, le navire ne comporte aucune de ses pièces originales.

Appelons B le navire qui est celui de Thésée lors de son retour en Grèce. Maintenant, la question est : le navire A est-il identique au navire B ? Autrement dit, le navire B est-il toujours le navire de Thésée ?

Il est intéressant de modéliser ce problème de manière plus précise. On peut considérer ainsi que le navire A possède *n* pièces (planches, pièces métalliques, cordes, etc.) qui sont autant de parties, qui peuvent être dénotées par $a_1$, $a_2$, $a_3$, ..., $a_n$, $a_n$. De même, les parties du navire B sont $b_1$, $b_2$, $b_3$, ..., $b_{n-1}$, $b_n$. On dénote ainsi le navire A par $a_1a_2a_3 \ldots a_{n-1}a_n$ et le navire B par $b_1b_2b_3 \ldots b_{n-1}b_n$. Au fil des années, c'est-à-dire du temps $T_0$ au temps $T_n$, le processus de remplacement des *n* pièces comporte les étapes successives suivantes :

| | | |
|---|---|---|
| (1) | $a_1a_2a_3 \ldots a_{n-1}a_n$ | en $T_0$ |
| (2) | $b_1a_2a_3 \ldots a_{n-1}a_n$ | en $T_1$ |
| (3) | $b_1b_2a_3 \ldots a_{n-1}a_n$ | en $T_2$ |
| (4) | $b_1b_2b_3 \ldots a_{n-1}a_n$ | en $T_3$ |
| (...) | | ... |
| (5) | $b_1b_2b_3 \ldots b_{n-1}a_n$ | en $T_{n-1}$ |
| (6) | $b_1b_2b_3 \ldots b_{n-1}b_n$ | en $T_n$ |

Il apparaît à ce stade que deux hypothèses peuvent être formulées :

(7) le navire B est identique au navire A
(8) le navire B n'est pas identique au navire A

De manière intuitive, ce qui justifie le fait que les navires A et B sont identiques, c'est que dans la vie courante, le simple fait de changer une pièce d'un appareil n'entraîne pas que cet appareil soit différent. De la même manière, intuitivement, l'identité du navire demeure identique à chaque fois qu'une planche ou une pièce métallique est remplacée. Sur ce fondement, on peut donc conclure que le navire B est identique au navire A.

Cependant, un autre argument plaide, de manière inverse, en faveur de l'hypothèse selon laquelle les navires A et B ne sont pas identiques. En effet, toutes les pièces du navire A ont été changées au fil des années. Ainsi, le navire B ne possède aucune des pièces originales du navire A. Comment, dans ces conditions, peut-on considérer que les navires A et B sont identiques ? En vertu du principe selon lequel deux objets qui ne possèdent aucune partie en commun sont distincts, la conclusion que les deux navires sont différents s'ensuit.

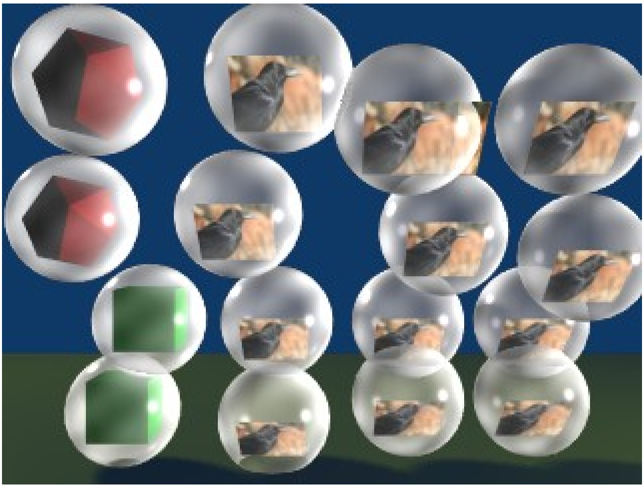
La description du problème du navire de Thésée s'accompagne souvent d'une seconde partie qui est la suivante. Alors que le navire s'éloigne de Grèce au moment du départ, il est accompagné d'un deuxième navire, chargé de l'assistance. A chaque fois qu'une réparation est effectuée sur le navire de Thésée, le navire d'assistance récupère l'ancienne pièce qui a été changée. Et le capitaine du navire d'assistance décide, à l'aide de son équipage, de reconstruire à l'identique le navire de Thésée original. De la sorte, lorsqu'il parvient en Grèce à son retour, ce second navire possède toutes les planches du navire original. Soit C le navire d'assistance. La question est alors : le navire C est-il identique au navire A ? Maintenant, il apparaît de manière encore plus nette que précédemment que le navire C est identique au navire A, puisque tous deux sont composés exactement des mêmes planches. Dans cette dernière version du problème du navire de Thésée, on a désormais quatre hypothèses :

(9) le navire B est identique au navire A et le navire C est identique au navire A
(10) le navire B est identique au navire A et le navire C n'est pas identique au navire A
(11) le navire B n'est pas identique au navire A et le navire C est identique au navire A
(12) le navire B n'est pas identique au navire A et le navire C n'est pas identique au navire A

Un premier type de solution qui a été proposé pour résoudre le problème du navire de Thésée repose sur l'idée qu'il ne s'agit que d'une variation du paradoxe sorite. Pourtant, un examen plus approfondi révèle que le problème du navire de Thésée est fondé sur la définition des critères de l'identité entre deux objets. La question cruciale qui apparaît ici est : dans quelles conditions un objet A est-il identique à un objet B ; et en particulier dans quelles conditions l'identité d'un objet persiste-t-elle à travers le temps ? En l'absence d'une réponse consensuelle à cette dernière question, on peut considérer que l'on ne dispose pas d'une solution satisfaisante pour le problème du navire de Thésée.

Un autre type de solution a été avancé par Derek Parfit, dans son ouvrage *Reasons and Persons* publié en 1984. Selon Parfit, c'est le fait de formuler les deux hypothèses en termes de relation d'*identité* qui se trouve à l'origine du problème. Car il faudrait reformuler le problème par rapport à un autre type de relation, qui peut être dénotée par R. Et il en résulte alors la conclusion selon laquelle le navire original de Thésée se trouve en relation R avec les deux navires, A et B. Pourtant, une telle analyse ne se révèle pas entièrement convaincante. Car le fait de remplacer la relation d'*identité* par une autre relation élimine en effet le problème. Mais une telle solution ne répond pas véritablement à la question pressante posée par le problème du navire de Thésée, qui porte précisément sur notre notion intuitive d'*identité* et les conditions de sa persistance temporelle.

## 17. Le problème de Hempel



Le *problème de Hempel* a été décrit par Carl Hempel, dans un article publié en 1945 dans la revue *Mind*, dans le cadre de l'étude de la théorie de la confirmation. Le point de départ en est l'assertion suivante : « tous les corbeaux sont noirs ». Clairement, la découverte d'un corbeau noir confirme une telle hypothèse. De même, cette hypothèse serait également infirmée par la découverte d'un corbeau bleu. Cependant, il s'avère que l'assertion selon laquelle « tous les corbeaux sont noirs » est équivalente à l'affirmation selon laquelle : « tout les objets non-noirs sont des non-corbeaux ». De même, on peut considérer valablement que tout ce qui confirme une proposition P donnée confirme également une proposition P* qui lui est équivalente. Mais ceci a alors pour conséquence que la découverte d'un flamand rose ou d'un parapluie bleu, qui confirme l'affirmation selon laquelle « tout les objets non-noirs sont des non-corbeaux », confirme également l'assertion selon

laquelle « tous les corbeaux sont noirs ». Et cette dernière conclusion apparaît paradoxale.

Le raisonnement sur lequel est basé le problème de Hempel peut être ainsi décrit de manière détaillée :

| | | |
|---|---|---|
| (1) | Tous les corbeaux sont noirs | *hypothèse 1* |
| (2) | Tout les objets non-noirs sont des non-corbeaux | *hypothèse 2* |
| (3) | (2) est équivalent à (1) | *contraposition* |
| (4) | les instances qui confirment une proposition P confirment également une proposition P* qui lui est équivalente | *prémisse* |
| (5) | la découverte d'un flamand rose confirme (2) | *de (3),(4)* |
| (6) | ∴ la découverte d'un flamand rose confirme (1) | *de (4),(5)* |

On peut observer ici que la structure logique de la proposition (1) selon laquelle « Tous les corbeaux sont noirs » présente la forme :

(7) Tous les X sont Y

alors que celle de (2) selon laquelle « Tout les objets non-noirs sont des non-corbeaux » est la suivante :

(8) Tous les non-Y sont non-X

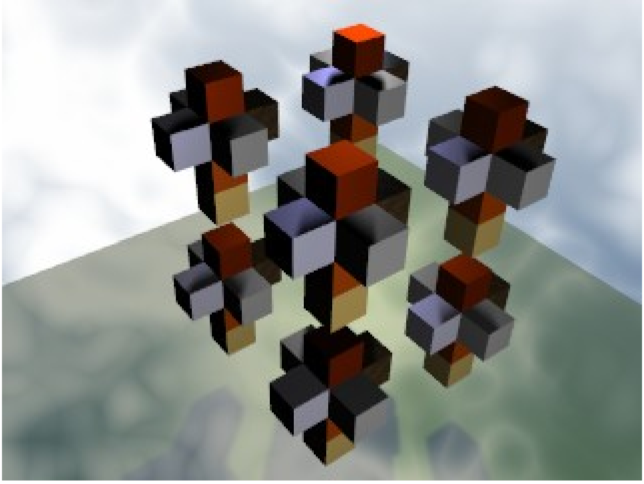
De fait, la structure de la forme contraposée (8) est clairement équivalente à celle de (7). On le voit, les propositions (1) et (2) sont basées sur quatre propriétés, qui correspondent respectivement à : *corbeau*, *non-corbeau*, *noir*, et *non-noir*. Ces quatre propriétés déterminent elles-mêmes quatre catégories d'objets : les *corbeaux noirs*, les *corbeaux non-noirs*, *les non-corbeaux noirs* et les *non-corbeaux non-noirs*.

On peut observer ici que le problème de Hempel ne constitue pas, au sens strict, un paradoxe. Car il n'en résulte pas une véritable contradiction. En revanche, la conclusion qui résulte du raisonnement inhérent au problème de Hempel se révèle fortement contraire à l'intuition. Pourtant, l'une des solutions qui a été proposée pour résoudre le problème de Hempel est basée sur l'acceptation de sa conclusion (6). Selon cette solution, la découverte d'un flamand rose confirme effectivement que tous les corbeaux sont noirs, mais seulement à un degré *infinitésimal*. Car la classe des non-corbeaux contient un nombre d'objets extrêmement élevé. Ainsi, selon ce type de solution, la découverte d'un non-corbeau confirme bien la proposition (1) selon laquelle « Tous les corbeaux sont noirs », mais seulement de manière infinitésimale.

Paul Feyerabend, dans un article publié en 1968 dans la revue *British Journal for the Philosophy of Science*, considère que le paradoxe de Hempel et celui de Goodman admettent un même type de solution. Selon Feyerabend, on ne doit considérer valablement, d'un point de vue scientifique, que les instances négatives (celles qui infirment une hypothèse), ce qui conduit à ignorer purement et simplement les instances positives (celles qui confirment une hypothèse). Dès lors que l'on ignore ces dernières, l'étape qui conduit à placer sur un même plan les instances confirmant (2) et celles confirmant (1) se trouve bloquée. Et dès lors, le paradoxe disparaît. Cependant, l'approche de Feyerabend s'est avérée trop radicale. Car il apparaît que confirmer une hypothèse H1, c'est également réfuter l'hypothèse inverse H2. Et réciproquement, réfuter l'hypothèse H1, c'est également confirmer l'hypothèse inverse H2. Ainsi, une instance donnée constitue une instance positive pour une hypothèse donnée en même temps qu'une instance négative pour l'hypothèse inverse. Pour cette raison, l'approche de Feyerabend n'est pas apparue véritablement convaincante.

Un autre type de solution qui a été proposé pour résoudre le problème de Hempel, est qu'un prédicat tel que « non-noir » ne devrait pas être utilisé sans restriction dans la pratique inductive. En effet, selon ce type de solution, il convient de se limiter aux prédicats qui sont projetables, car tout prédicat est susceptible de donner lieu à nombreuses variations construites sur le modèle de « vleu ». Selon ce type d'analyse, le problème de Hempel et le paradoxe de Goodman sont le résultat de l'application sans restriction de tous les prédicat dans les processus inductifs. Pourtant, une telle analyse ne se révèle pas non plus convaincante. En effet, « non-noir », à la différence de « vleu », ne comporte pas de clause temporelle. Et c'est ici « non-noir » qui est projeté, et non pas « non-noir avant T ». Et renoncer à toute projection inductive d'un prédicat présentant la structure « non-P » constitue un sacrifice trop important pour résoudre le paradoxe.

## 18. L'argument de McTaggart



Dans un article resté célèbre, publié en 1908 dans la revue *Mind*, John Ellis McTaggart a décrit un argument destiné à prouver que le temps n'est pas réel. McTaggart commence par distinguer deux types de propriétés des positions temporelles :

> Les positions temporelles, ainsi que le temps nous apparaît à première vue, peuvent être distinguées de deux façons. Chaque position temporelle se trouve *avant* certaines autres et *après* d'autres positions... En second lieu, chaque position temporelle est soit *passée*, *présente* ou *future*. Les distinctions de la première classe sont permanentes, alors que celles de la seconde classe ne le sont pas. Si un événement M a lieu avant un autre événement N, alors il se trouve toujours placé avant ; mais un événement, qui est maintenant présent, a été futur, et sera passé.

McTaggart appelle série B la première distinction, en vertu de laquelle toute position temporelle M est placée *avant* mais aussi *après* d'autres positions temporelles. Il mentionne également une propriété constante des séries B : lorsqu'un événement M est antérieur a un événement N à un moment donné, il se révèle être antérieur à N de manière permanente. McTaggart dénomme également série A la seconde distinction, en vertu de laquelle toute position temporelle M appartient soit au *passé*, soit au *présent*, soit au *futur*. McTaggart observe que les séries A sont telles que chaque événement M est tour à tour passé, présent et futur. Ainsi, un événement qui est présent, a été futur et sera passé. De même, un événement qui est passé, a été présent et futur. Enfin, un événement qui est futur, sera présent et passé. Ainsi, la seconde distinction met en évidence un élément non permanent au niveau du temps.

McTaggart poursuit ensuite son raisonnement en montrant comment le temps doit nécessairement présenter toutes les propriétés des séries A. Car supposons que le temps soit défini uniquement à l'aide des séries B. Dans ce cas, on ne peut rendre compte d'un élément essentiel du temps, à savoir le changement. Ainsi, poursuit McTaggart, il s'avère nécessaire de recourir aux séries A pour rendre compte des propriétés essentielles du temps.

Enfin, McTaggart s'attache à démontrer comment les propriétés des séries A conduisent à une contradiction. Car les séries A sont mutuellement exclusives : un événement ne peut être à la fois passé, présent et futur. L'intuition qui préside à notre notion de temps est qu'un événement donné ne peut être passé, présent et futur *simultanément*. Pourtant, McTaggart considère une position temporelle donnée M : cette dernière est présente, sera passée et a été future. Mais « sera passé » équivaut à « est passé à une position temporelle future » ; et de même, « a été futur » équivaut à « est futur à une position temporelle passée ». Ainsi, on définit *passé* par rapport à *futur*, et *futur* par rapport à *passé*. Il en résulte donc une définition circulaire. Ceci montre l'incohérence des séries A. Par conséquent, aucun événement ne peut posséder toutes les propriétés des séries A. Il s'ensuit que le temps ne peut présenter toutes les propriétés des séries A. Ainsi, conclut McTaggart, le temps ne possède pas de réalité.

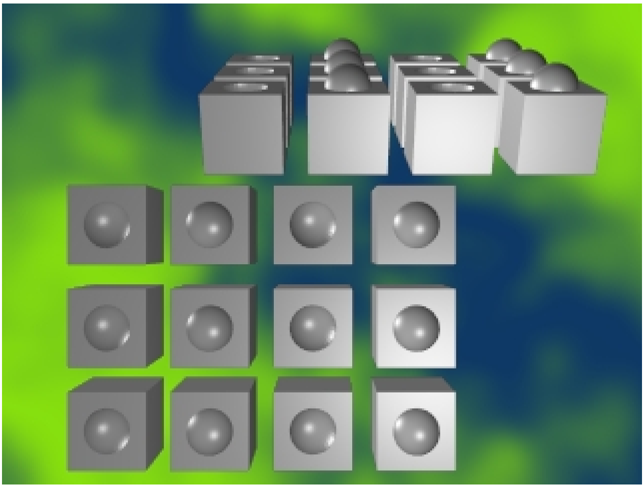
La structure de l'argument de McTaggart peut ainsi être mise en évidence de manière détaillée de la façon suivante :

| | | |
|---|---|---|
| (1) | toute position temporelle possède deux propriétés distinctes : la série A et la série B | *prémisse* |
| (2) | la série B ne permet pas de rendre compte du changement | *prémisse* |
| (3) | le changement est un élément essentiel du temps | *prémisse* |
| (4) | la série B ne permet pas de rendre compte d'un élément essentiel du temps | *de (2),(3)* |
| (5) | le temps doit posséder les propriétés de la série A pour rendre compte d'un élément essentiel, le changement | *de (1),(4)* |
| (6) | le temps possède les propriétés de la série A | *hypothèse* |
| (7) | dans la série A, un événement futur est défini par rapport au passé | |
| (8) | dans la série A, un événement présent est défini par rapport au présent | |
| (9) | dans la série A, un événement passé est défini par rapport au présent | |
| (10) | ∴ dans la série A, les définitions sont circulaires | *de (7),(8), (9)* |
| (11) | le temps ne peut posséder les propriétés de la série A | *de (10)* |
| (12) | ∴ le temps est irréel | *de (5), (11)* |

Une objection qui peut être opposée à l'argument de McTaggart est que le fait que les séries B ne suffisent pas à rendre compte des propriétés essentielles du temps ne prouve pas qu'il est indispensable de recourir aux séries A. Car peut-être pourrait-on trouver une autre série – appelons-la série D – qui permettrait de rendre compte des propriétés du temps, en combinaison avec les séries B, mais sans présenter les inconvénients des séries A. En d'autres termes, il existe peut-être d'autres alternatives aux séries A, qui permettraient de rendre compte de manière adéquate des propriétés intrinsèques du temps.

Une autre objection qui a été formulée, à l'encontre de l'argument de McTaggart, en particulier par Bertrand Russell, est que les séries A peuvent être obtenues logiquement à partir des séries B. Ainsi, selon Russell, les notions de passé, présent, futur peuvent être définies à partir des relations *avant*, *pendant*, *après*, qui constituent alors les termes primitifs. Ainsi, *passé*, *présent*, *futur* sont respectivement définis comme : avant T, pendant T, après T. L'objection de Russell a pour but de montrer comment les séries A ne sont finalement pas nécessaires pour décrire les propriétés du temps. Cependant, la définition de Russell présente l'inconvénient de comporter une référence au moment T. Et on peut penser que cette référence implicite à T s'assimile au « moment présent ». Ceci conduit finalement à définir le présent comme « pendant le moment présent », d'une manière qui s'avère toutefois également circulaire.

# 19. L'argument ontologique



Un argument ontologique est un argument qui conclut à l'existence de Dieu, à partir de considérations *a priori*, c'est-à-dire de prémisses qui ne sont pas basées sur des constatations empiriques ou des preuves matérielles. Un argument ontologique a pour objet de constituer une preuve de l'existence de Dieu. Cependant, à la différence des preuves classiques qui résultent de l'observation du réel, une telle preuve est basée uniquement sur le raisonnement. Il existe ainsi plusieurs types d'arguments ontologiques. Le plus ancien est dû à Saint Anselme de Canterbury (1077). Le point de départ en est la prise en considération d'un être dont on ne peut pas concevoir un être plus grand. Si celui-ci n'existe pas, on peut dès lors concevoir un être dont on ne peut concevoir un être plus grand et qui de surcroît existe. Mais ceci implique que l'on peut concevoir un être plus grand que l'être dont on ne peut concevoir un être plus grand. Et cette dernière conclusion se révèle contradictoire. Ainsi, la prise en compte de l'hypothèse

selon laquelle l'être dont on ne peut concevoir un être plus grand n'existe pas, conduit à une contradiction. Par conséquent, l'être dont on ne peut concevoir un être plus grand existe. L'argument ontologique de Saint Anselme peut être décrit ainsi de manière détaillée :

| | | |
|---|---|---|
| (1) | je peux concevoir un être dont on ne peut concevoir un être plus grand | *prémisse* |
| (2) | soit l'être dont on ne peut concevoir un être plus grand existe, soit il n'existe pas | *dichotomie* |
| (3) | si un être dont on ne peut concevoir un être plus grand n'existe pas | *hypothèse 1* |
| (4) | alors je peux concevoir un être dont on ne peut concevoir un être plus grand mais qui existe | *de (3)* |
| (5) | je peux concevoir un être plus grand que l'être dont on ne peut concevoir un être plus grand | *de (3),(4)* |
| (6) | ∴ un être dont on ne peut concevoir un être plus grand existe | *de (2),(4)* |

Un argument ontologique légèrement différent est dû à Descartes, qui le décrit ainsi dans ses *Méditations*. Selon Descartes, Dieu, par définition, est un être parfait. Il possède donc toutes les qualités. Par conséquent, il possède également celle d'exister. Dieu existe donc. L'argument ontologique de Descartes met l'accent sur la définition de Dieu en tant qu'être parfait. Le passage original des *Méditations* qui contient l'argument ontologique de Descartes est le suivant :

> Or maintenant, si de cela seul que je puis tirer de ma pensée l'idée de quelque chose, il s'ensuit que tout ce que je reconnais clairement et distinctement appartenir à cette chose, lui appartient en effet, ne puis-je pas tirer de ceci

> un argument et une preuve démonstrative de l'existence de Dieu ? Il est certain que je ne trouve pas moins en moi son idée, c'est-à-dire l'idée d'un être souverainement parfait, que celle de quelque figure ou de quelque nombre que ce soit. Et je ne connais pas moins clairement et distinctement qu'une actuelle et éternelle existence appartient à sa nature, que je connais que tout ce que je puis démontrer de quelque figure ou de quelque nombre, appartient véritablement à la nature de cette figure ou de ce nombre. Et partant, encore que tout ce que j'ai conclu dans les Méditations précédentes, ne se trouvât point véritable, l'existence de Dieu doit passer en mon esprit au moins pour aussi certaine, que j'ai estimé jusqu'ici toutes les vérités des mathématiques, qui ne regardent que les nombres et les figures : bien qu'à la vérité, cela ne paraisse pas d'abord entièrement manifeste, mais semble avoir quelque apparence de sophisme. Car, ayant accoutumé dans toutes les autres choses de faire distinction entre l'existence et l'essence, je me persuade aisément que l'existence peut être séparée de l'essence de Dieu, et qu'ainsi on peut concevoir Dieu comme n'étant pas actuellement. Mais néanmoins, lorsque j'y pense avec plus d'attention, je trouve manifestement que l'existence ne peut non plus être séparée de l'essence de Dieu, que de l'essence d'un triangle rectiligne la grandeur de ses trois angles égaux à deux droits, ou bien de l'idée d'une montagne l'idée d'une vallée ; en sorte qu'il n'y a pas moins de répugnance de concevoir un Dieu (c'est-à-dire un être souverainement parfait) auquel manque l'existence (c'est-à-dire auquel manque quelque perfection), que de concevoir une montagne qui n'ait point de vallée.

De manière plus précise, la structure de l'argument ontologique de Descartes peut être ainsi définie :

| | | |
|---|---|---|
| (1) | Dieu est un être parfait | *définition* |
| (2) | Dieu est un être qui possède toutes les qualités | *de (1)* |
| (3) | l'existence constitue une qualité | *prémisse* |

| | | |
|---|---|---|
| (4) | ∴ Dieu existe | *de (2),(3)* |

Les arguments ontologiques ont fait l'objet, dans la littérature, de multiples objections. Une critique célèbre émane notamment de Kant, dans sa *Critique de la raison pure*, qui considère que l'existence ne constitue pas une authentique propriété. Ceci a pour conséquence de bloquer la prémisse (3) de l'argument ontologique de Descartes, neutralisant ainsi le raisonnement qui conduit à la conclusion selon laquelle Dieu existe. Selon Kant, on ne peut considérer que le fait d'exister constitue une propriété, au même titre que *rouge* constitue la propriété d'une tomate, ou *dur* constitue la propriété d'une pierre. Pour Kant, c'est l'existence même d'une chose *x* qui constitue une condition nécessaire pour l'attribution de ses propriétés (couleur, dimensions, densité, rugosité, dureté, etc.).

D'une manière générale, les arguments ontologiques ne sont habituellement pas considérés comme des preuves véritablement convaincantes de l'existence de Dieu et ils se révèlent en général insuffisants pour convaincre des non-théistes de l'existence de Dieu.

## 20. L'argument du réglage optimal

L'*argument du réglage optimal* (*fine-tuning argument*) appartient à la catégorie des arguments qui visent à démontrer l'existence de Dieu. L'argument repose sur le fait qu'un nombre important de constantes cosmologiques régissant notre univers sont telles que si elles avaient été très légèrement différentes, l'émergence de la vie intelligente basée sur la chimie du carbone telle que nous l'observons sur Terre n'aurait pas été possible. Parmi ces constantes, on peut citer : le rapport des masses respectives de l'électron et du proton, l'âge de l'univers, la masse du neutrino, la distance moyenne entre les étoiles, la vitesse de la lumière, la constante cosmologique universelle, la constante de Planck, etc. L'argument est sous-tendu par le fait que chacun de ces paramètres aurait pu avoir une valeur légèrement différente, mais qui n'aurait pas permis alors l'émergence de la vie. Considérons par exemple la vitesse de la lumière dans le vide ($v$ = 299792,458 km/s) : si celle-ci avait été ne serait-ce que légèrement plus élevée, les étoiles auraient émis trop de lumière pour permettre l'émergence de la vie ; et de même, si la vitesse de la lumière avait été à peine plus faible, l'émission de lumière par les étoiles aurait été insuffisante pour permettre l'apparition de la vie. Il en va de même pour la constante gravitationnelle ($G = 6{,}672.10^{-11}\ Nm^2kg^{-2}$) : si cette dernière avait eu une valeur légèrement plus élevée, les étoiles auraient eu une température trop haute et se seraient consumées beaucoup trop vite pour permettre l'émergence de la vie basée sur la chimie du carbone. De même, si la constante gravitationnelle avait été légèrement plus faible, la température des étoiles auraient été trop basse pour permettre la formation de nombreux éléments chimiques nécessaires à l'apparition de la vie. On peut considérer également le ratio de la masse de l'électron par rapport à celle du proton ($m_e/m_p = 5.446170232x10^{-4}$) : s'il avait été légèrement différent, les liaisons chimiques qui en auraient résulté auraient été insuffisantes pour permettre l'apparition de la vie. Enfin, si le taux de l'expansion de l'univers avait été à légèrement supérieur, aucune galaxie n'aurait pu se former ; et de même, s'il avait été à légèrement inférieur, l'univers se serait effondré, avant même la formation des étoiles, etc.

Ainsi, ces différents paramètres, en vertu de l'argument du réglage optimal, n'ont pas été déterminés au hasard, mais en fonction d'une finalité particulière : l'apparition de la vie intelligente dans l'univers. Ce but particulier témoigne de la présence d'un dessein divin et donc finalement de l'existence de Dieu.

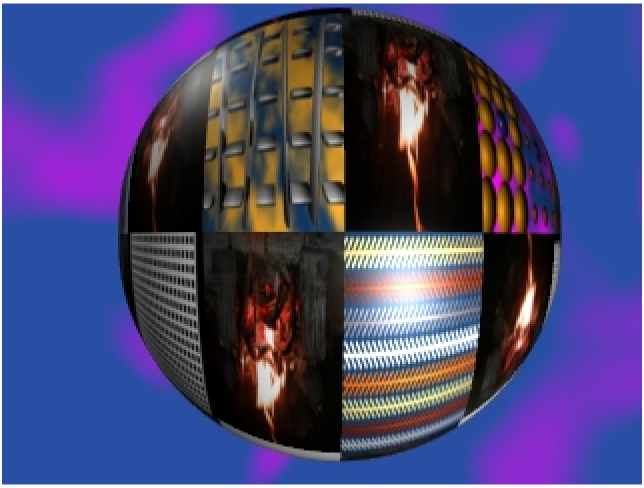
On peut détailler ainsi les différentes étapes de l'argument du réglage optimal :

| | | |
|---|---|---|
| (1) | plusieurs constantes cosmologiques régissant notre univers ont des valeurs telles qu'elles permettent l'émergence de la vie intelligente basée sur la chimie du carbone | *prémisse* |

| | | |
|---|---|---|
| (2) | les constantes cosmologiques régissant notre univers auraient pu avoir un grand nombre de valeurs différentes | *hypothèse* |
| (3) | si les valeurs de ces constantes cosmologiques avaient été légèrement différentes, alors l'émergence de la vie intelligente basée sur la chimie du carbone n'aurait pas été possible | *de (2)* |
| (4) | si les constantes cosmologiques avaient été obtenues au hasard, alors la probabilité que leur réglage soit optimal aurait été extrêmement faible | *de (2)* |
| (5) | le réglage optimal des constantes cosmologiques ne résulte pas du hasard | *de (4)* |
| (6) | le réglage optimal des constantes cosmologiques a été effectué à dessein afin de permettre l'émergence de la vie intelligente | *de (3),(5)* |
| (7) | le réglage optimal des constantes cosmologiques a été effectué par Dieu | *de (6)* |
| (8) | ∴ Dieu existe | *de (7)* |

Plusieurs objections ont été opposées à l'argument du réglage optimal. L'une d'entre elles en particulier repose sur l'idée spéculative, défendue par un certain nombre de cosmologistes, selon laquelle l'univers que nous observons n'est pas le seul, mais constitue seulement un univers parmi de très nombreux autres, au sein d'un système composé de multiples univers causalement indépendants. Dans ce contexte, il existe un grand nombre d'autres univers, complètement différents du notre, qui possèdent des paramètres cosmologiques tout à fait distincts. On le voit, cette objection vise directement l'étape (5) du raisonnement qui sous-tend l'argument du réglage optimal, selon laquelle le réglage optimal des constantes cosmologiques ne résulte pas du hasard. Car l'hypothèse des univers multiples s'avère tout à fait compatible avec le fait que les paramètres de notre univers puissent avoir été obtenus de manière aléatoire.

## 21. L'argument du rêve



L'argument du rêve est dû à Descartes. Il peut être formulé très simplement. Il s'agit d'un argument qui conduit à la conclusion que nos perceptions actuelles pourraient bien être illusoires et trompeuses, car elles sont en tous points analogues à celles que nous avons lorsque nous rêvons. Lorsque nous sommes en effet en état de rêve, nos perceptions sont en effet suffisamment réalistes pour être capables de créer l'illusion de la réalité. L'argument du rêve est décrit dans le passage suivant (*Première méditation*) des *Méditations métaphysiques* :

> Mais, encore que les sens nous trompent quelquefois, touchant les choses peu sensibles et fort éloignées, il s'en rencontre peut-être beaucoup d'autres, desquelles on ne peut pas raisonnablement douter, quoique nous les connaissions par leur moyen : par exemple, que je sois ici, assis auprès du feu, vêtu d'une robe de chambre, ayant ce papier entre les mains, et autres choses de cette nature. Et comment est-ce que je pourrais nier que ces

> mains et ce corps-ci soient à moi ? Si ce n'est peut-être que je me compare à ces insensés, de qui le cerveau est tellement troublé et offusqué par les noires vapeurs de la bile, qu'ils assurent constamment qu'ils sont des rois, lorsqu'ils sont très pauvres ; qu'ils sont vêtus d'or et de pourpre, lorsqu'ils sont tout nus ; ou s'imaginent être des cruches, ou avoir un corps de verre. Mais quoi ? Ce sont des fous, et je ne serais pas moins extravagant, si je me réglais sur leurs exemples.
>
> Toutefois j'ai ici à considérer que je suis homme, et par conséquent que j'ai coutume de dormir et de me représenter en mes songes les mêmes choses, ou quelquefois de moins vraisemblables, que ces insensés, lorsqu'ils veillent. Combien de fois m'est-il arrivé de songer, la nuit, que j'étais en ce lieu, que j'étais habillé, que j'étais auprès du feu, quoique je fusse tout nu dedans mon lit ? Il me semble bien à présent que ce n'est point avec des yeux endormis que je regarde ce papier ; que cette tête que le remue n'est point assoupie ; que c'est avec dessein et de propos délibéré que j'étends cette main, et que je la sens : ce qui arrive dans le sommeil ne semble point si clair ni si distinct que tout ceci. Mais, en y pensant soigneusement, je me ressouviens d'avoir été souvent trompé, lorsque je dormais, par de semblables illusions. Et m'arrêtant sur cette pensée, je vois si manifestement qu'il n'y a point d'indices concluants, ni de marques assez certaines par où l'on puisse distinguer nettement la veille d'avec le sommeil, que j'en suis tout étonné ; et mon étonnement est tel, qu'il est presque capable de me persuader que je dors.

L'argument du rêve peut être détaillé de la manière suivante :

| | | |
|---|---|---|
| (1) | lorsque je suis éveillé, j'ai des perceptions | *prémisse* |
| (2) | lorsque je rêve, j'ai également des perceptions | *prémisse* |
| (3) | les perceptions que j'ai lorsque je suis éveillé sont en tous points identiques à | *prémisse* |

| | | |
|---|---|---|
| | celles que j'ai lorsque je rêve | |
| (4) | je ne possède pas de critère qui me permette de distinguer mes perceptions lorsque je suis éveillé ou lorsque je rêve | *de (3)* |
| (5) | je n'ai pas de preuve que je ne suis pas actuellement en état de rêve | *de (4)* |
| (6) | ∴ il est possible que je sois actuellement en état de rêve | *de (5)* |
| (7) | lorsque je rêve, mes perceptions sont fausses | *prémisse* |
| (8) | ∴ il est possible que toutes mes perceptions actuelles soient fausses | *de (6),(7)* |

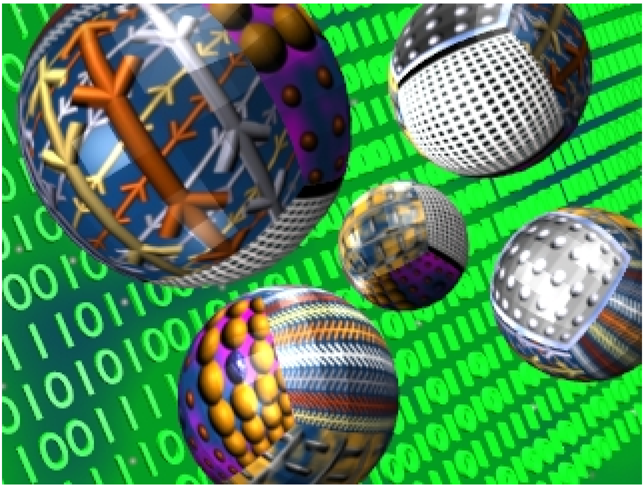
L'argument du rêve de Descartes a donné lieu à plusieurs variations contemporaines. L'une de ces variations modernes repose sur l'idée que nous sommes des « cerveaux dans une cuve ». Le film *Matrix*, de Larry et Andy Wachowski développe également une variante de cette idée.

Une objection à l'argument du rêve a été soulevée par Barry Stroud, dans un ouvrage paru en 1989. Selon cette objection, la prémisse (4) s'avère fausse, car il est tout à fait possible d'effectuer un test qui permet de déterminer si chacun d'entre nous est ou non en état de rêve. A l'aide de capteurs qui déterminent si les ondes cérébrales caractéristiques de l'état de rêve sont produites par le cerveau, on peut déterminer si une personne rêve ou non, et apporter ainsi une réponse définitive et fiable à cette question. Cependant, cette objection n'a pas convaincu plusieurs auteurs,

qui ont fait valoir qu'une telle réponse présuppose que l'on ne rêve pas au moment où on effectue le test. Dans cette hypothèse, le fait d'effectuer un test se révèle effectivement concluant. Mais supposons à l'inverse que nous soyons en état de rêve au moment où nous effectuons le test. Dans ce cas, le test fait partie de notre rêve et on ne peut valablement lui accorder notre confiance. Ainsi, l'idée qui sous-tend cette objection présuppose finalement que nous ne rêvons pas, alors que précisément, c'est cette question-même qui est véritablement posée.

Un autre type d'objection peut également être soulevé par rapport à l'argument du rêve. Supposons que ce dernier argument soit tout à fait valide et que sa conclusion soit irréfutable. Dans ce cas, on dispose alors d'une preuve inébranlable que nous sommes en état de rêve. Mais si tel était le cas, ne s'ensuivrait-il pas alors que l'argument du rêve lui-même n'est qu'un pur produit de notre rêve, et donc quelque chose d'illusoire. Ainsi, en aucun cas il ne pourrait s'agir d'un raisonnement sur lequel nous pourrions baser nos connaissances. On le voit, une telle propriété a pour effet de rendre l'argument du rêve auto-réfutant.

## 22. L'expérience des « cerveaux dans une cuve »




L'expérience des « cerveaux dans une cuve » a été énoncée par Hilary Putnam, dans son ouvrage *Raison, Vérité et Histoire* paru en 1982. L'argument commence par l'interrogation suivante : est-ce que je ne suis pas un cerveau dans une cuve ? Autrement dit, suis-je bien certain que quelque savant fou ne m'a pas enlevé, n'a pas ensuite prélevé mon cerveau pour le placer dans un liquide nutritif, et n'a pas enfin simulé toutes les informations qui parviennent d'habitude à mon cerveau, à l'aide d'une dispositif particulièrement sophistiqué. De la sorte, mes sensations, mes perceptions, mes pensées, etc. ne seraient que l'effet des stimulations que le savant fou envoie, à l'aide de son appareillage, à l'ensemble de mes neurones. Suis-je bien tout à fait certain que je ne me trouve pas dans une situation de ce type ? Si tel était le cas, les stimulations envoyées à mon cerveau seraient telles qu'elles produiraient exactement les impressions qui sont les miennes lorsque j'ai des sensations, des perceptions, des émotions ou des pensées, dans des conditions normales. Comment donc puis-je être tout à fait certain que je ne suis pas un cerveau dans une cuve ?

Cependant, l'argument de Putnam n'a pas pour finalité de suggérer que nous sommes réellement des cerveaux dans des cuves. Pour Putnam, il est en effet clair, au contraire, que nous ne sommes pas des « cerveaux dans des cuves ». Pour lui, ceci résulte de la simple considération de l'assertion selon laquelle « nous sommes des cerveaux dans des cuves ». Putnam se propose de prouver que cette dernière assertion est toujours fausse. Il distingue ainsi deux hypothèses : si (a) nous ne sommes pas des cerveaux dans des cuves, alors il est faux que nous sommes des cerveaux dans des cuves ; si (b) nous sommes des cerveaux dans des cuves, alors les concepts et les mots que nous utilisons quotidiennement se réfèrent non à des objets réels, mais à des objets virtuels, qui sont le résultat d'une simulation. Tel est le cas lorsque nous utilisons des concepts tels que « table », « chaise », « parapluie », etc. Dans ce cas, nos concepts de « table » ou « parapluie » se réfèrent non pas à une table ou un parapluie, mais à une simulation de table ou de parapluie qui provient des impulsions électriques envoyées à nos cerveaux par un dispositif électronique sophistiqué. Et tel est également le cas lorsque nous faisons usage de mots tels que « cerveau » ou « cuve ». Dans ce cas, nous nous référons alors à une simulation de cerveau ou de cuve. Ainsi, lorsque nous affirmons que « nous sommes des cerveaux dans des cuves », nous énonçons le fait que « nous sommes des simulations de cerveaux dans des simulations de cuves ». Mais ceci ne correspond pas alors à la réalité. Ainsi, si l'on envisage l'hypothèse que nous sommes des cerveaux dans des cuves, il s'avère également qu'il est faux que nous sommes des cerveaux dans des cuves. En conclusion, quelle que soit l'hypothèse envisagée, il est faux que nous soyons des cerveaux dans des cuves.

L'argument de Putnam peut être décrit plus précisément de la manière suivante :

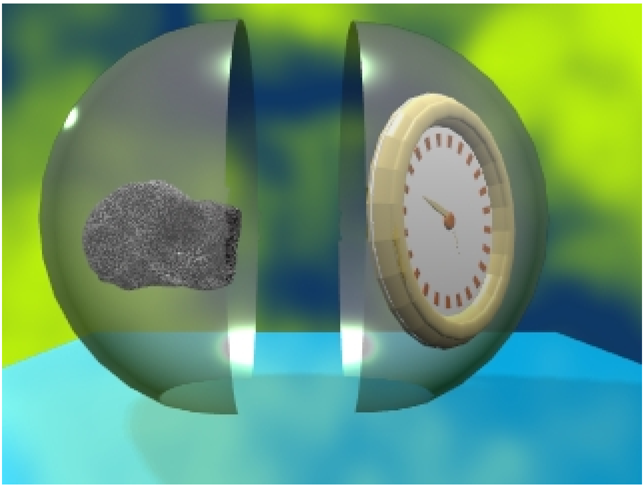
| | | |
|---|---|---|
| (1) | il n'existe pas de critère interne permettant de savoir si nos sensations, nos perceptions, nos émotions et nos pensées sont stimulées ou non par un dispositif | *prémisse* |
| (2) | si nous sommes des cerveaux dans des cuves | *hypothèse 1* |
| (3) | alors nos sensations, nos perceptions, nos émotions et nos pensées sont stimulées par un dispositif | *de (2)* |
| (4) | alors « nous sommes des cerveaux dans des cuves » signifie que « nous sommes des simulations de cerveaux dans des simulations de cuves » | *de (3)* |
| (5) | si nous ne sommes pas des cerveaux dans des cuves | *hypothèse 2* |
| (6) | alors nos sensations, nos perceptions, nos émotions et nos pensées ne sont pas stimulées par un dispositif | *de (5)* |
| (7) | alors il est faux que « nous sommes des cerveaux dans des cuves » | *de (5)* |
| (8) | ∴ il est nécessaire de recourir à un critère externe pour déterminer si nous sommes ou non des cerveaux dans des cuves | *de (1),(4), (7)* |

L'expérience de pensée de Putnam a pour but de souligner que les états internes qui résultent de la stimulation par un dispositif extérieur d'un cerveau dans une cuve d'une part, et les pensées et les perceptions d'une personne normale d'autre part, ne peuvent être distingués. Car les états mentaux internes qui en résultent sont dans les deux cas identiques. Par conséquent, il est nécessaire de recourir à des critères externes pour les différencier.

Ainsi le point de vue émis par Putnam se révèle-t-il fondamentalement *externaliste*. L'argument de Putnam souligne ainsi que la signification des mots ou des phrases ne dépend pas uniquement du contenu interne, c'est-à-dire de nos pensées, de nos émotions, etc. Selon la formule célèbre de Putnam, « le sens n'est pas seulement dans notre tête » [4].

[4] "Meanings just are not in the head".

# 23. L'argument téléologique



L'argument téléologique ou argument du dessein divin appartient, de même que l'argument ontologique, à une famille d'arguments qui visent à prouver l'existence de Dieu. L'argument téléologique repose sur l'idée simple que notre univers est si complexe et si bien agencé que cela ne peut être que la manifestation du dessein d'une entité intelligente. L'ordonnancement complexe de notre univers démontre ainsi que ce dernier possède un Créateur.

L'argument du dessein divin peut être décrit ainsi de manière plus détaillée :

| | | |
|---|---|---|
| (1) | notre univers est très complexe et très bien agencé | *prémisse* |
| (2) | la complexité et l'agencement de notre univers ne peut qu'être que la manifestation du dessein d'un être intelligent | *de (1)* |
| (3) | un être intelligent est le Créateur de | *de (2)* |

| | | |
|---|---|---|
| | notre univers | |
| (4) | ∴ Dieu est le Créateur de notre univers | *de (3)* |

Une formulation célèbre de l'argument du dessein divin est notamment due à William Paley (1743-1805), dans son ouvrage *Théologie naturelle* (*Natural Theology*), paru en 1802. Paley décrit l'argument dans les termes suivants :

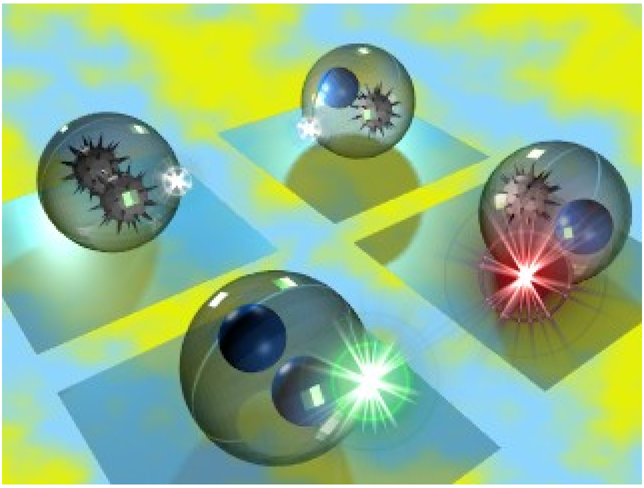
> En traversant une lande, supposons que je heurte du pied une *pierre* et que l'on demande pour quelle raison la pierre se trouvait là. Je pourrais alors peut-être répondre que pour autant que je sache, en l'absence d'information contraire, elle se trouvait là depuis toujours ; à ce stade, il ne devrait pas être très aisé de démontrer l'absurdité d'une telle réponse. Mais supposons maintenant que j'ai trouvé par terre une *montre*, et que l'on fasse une investigation pour savoir pour quelle raison la montre se trouvait à cet endroit précis. Dans ce cas, je pourrais difficilement faire appel à la raison donnée précédemment, à savoir que la montre s'était toujours trouvée là.

Selon Paley, la raison pour laquelle on ne peut concevoir que la montre se soit trouvée là depuis toujours, est que ses différentes parties ont été assemblées *à dessein*, et que ce dessein ne peut qu'être l'œuvre d'un être intelligent. L'argument de Paley est basé sur une analogie entre la montre et l'univers, et conduit à la conclusion que l'univers n'a pu qu'être créé qu'à dessein, et que ce dessein est celui de Dieu.

Une objection qui a été formulée contre l'argument du dessein divin est dirigée contre la prémisse (1), selon laquelle notre univers est très bien agencé. Mais à cela, il peut être rétorqué qu'il ne s'agit que de l'expression d'un point de vue spécifique concernant notre univers. Car d'un autre point de vue, notre univers pourrait apparaître comme très mal agencé. Il suffirait pour cela de considérer que le désordre est partout présent dans le monde. Car, pourrait-on faire observer, notre monde est agité par de fréquents tremblements de terre, raz-de-marée dévastateurs, cyclones destructeurs, etc. et subit, de manière générale, de nombreuses catastrophes naturelles. De ce point de vue, on ne peut véritablement considérer l'univers comme bien agencé.

Une autre objection vise directement l'étape (2) selon laquelle le bel ordonnancement de notre univers ne peut qu'être l'œuvre d'un être intelligent. En vertu de cette objection, la complexité de notre univers et son agencement sophistiqués sont bien avérés, mais cela n'implique pas pour autant que ceci soit l'œuvre d'un créateur. Car on pourrait également imaginer que de nombreux univers coexistent, certains étant très simples et rudimentaires, alors que d'autres sont complexes et sophistiqués. En tant qu'observateurs, nous ne pouvons évidemment nous trouver que dans un univers complexe et ordonnancé, permettant notamment l'émergence de la vie basée sur la chimie du carbone. Par contre, il pourrait tout à fait exister de nombreux univers très différents du notre, dont certains seraient très frustres et rudimentaires, et dépourvus d'observateurs.

## 24. L'argument du pari de Pascal



Le pari de Pascal est un argument contenu dans le paragraphe 233 des *Pensées*. Il s'agit là d'un des arguments les plus célèbres de la philosophie de la religion, qui se propose de fournir au lecteur de solides raisons de croire en l'existence de Dieu. Pascal y expose l'alternative devant laquelle nous trouvons placés : soit Dieu existe, soit il n'existe pas. Confrontés à une telle situation, nous pouvons donc parier en faveur de l'existence de Dieu, ou bien en faveur de sa non-existence. Pascal analyse ensuite les conséquences qui découlent d'un pari en faveur de l'une ou l'autre option. Il envisage ensuite les quatre cas qui sont ainsi déterminés. Si je parie en faveur de l'existence de Dieu et que Dieu existe (a), alors j'obtiens un gain infini. Si je parie en faveur de l'existence de Dieu et que Dieu n'existe pas (b), alors il en résulte une perte nulle. Si je parie pour la non-existence de Dieu et que Dieu existe (c), alors il s'ensuit une perte infinie. Enfin, si je parie pour la non-existence de Dieu et que Dieu n'existe pas (d), alors je n'obtiens ni gain ni perte. Ainsi, il apparaît que si je parie pour la non-existence de Dieu, je me trouve exposé à un perte infinie. Par conséquent, conclut Pascal, il est plus sage de parier en faveur de l'existence de Dieu, car il s'ensuit soit un gain infini, soit une perte nulle.

L'extrait des *Pensées* de Pascal qui contient l'argument du pari est le suivant :

> Je ne me servirai pas, pour vous convaincre de son existence, de la foi par laquelle nous la connaissons certainement, ni de toutes les autres preuves que nous en avons, puisque vous ne les voulez pas recevoir. Je ne veux agir avec vous que par vos principes mêmes ; et je ne prétends vous faire voir par la manière dont vous raisonnez tous les jours sur les choses de la moindre conséquence, de quelle sorte vous devez raisonner en celle-ci, et quel parti vous devez prendre dans la décision de cette importante question de l'existence de Dieu. Vous dites donc que nous sommes incapables de connaître s'il y a un Dieu. Cependant il est certain que Dieu est, ou qu'il n'est pas ; il n'y a point de milieu. Mais de quel côté pencherons- nous ? La raison, dites vous, n'y peut rien déterminer. Il y a un chaos infini qui nous sépare. Il se joue un jeu à cette distance infinie, où il arrivera croix ou pile. Que gagnerez vous ? Par raison vous ne pouvez assurer ni l'un ni l'autre ; par raison vous ne pouvez nier aucun des deux.
>
> Ne blâmez donc pas de fausseté ceux qui ont fait un choix ; car vous ne savez pas s'ils ont tort, et s'ils ont mal choisi. Non, direz vous ; mais je les blâmerai d'avoir fait non ce choix, mais un choix : et celui qui prend croix, et celui qui prend pile ont tous deux tort : le juste est de ne point parier.
>
> Oui ; mais il faut parier ; cela n'est pas volontaire ; vous êtes embarqué ; et ne parier point que Dieu est, c'est parier qu'il n'est pas. Lequel prendrez vous donc ? Pesons le gain et la perte en prenant le parti de croire que Dieu est. Si vous gagnez, vous gagnez tout ; si vous perdez, vous ne perdez rien. Pariez donc qu'il est sans hésiter. Oui il faut gager. Mais je gage peut-être trop. Voyons : puisqu'il y a pareil hasard de gain et de perte, quand vous

n'auriez que deux vies à gagner pour une, vous pourriez encore gager. Et s'il y en avait dix à gagner, vous seriez bien imprudent de ne pas hasarder votre vie pour en gagner dix à un jeu où il y a pareil hasard de perte et de gain. Mais il y a ici une infinité de vies infiniment heureuses à gagner avec pareil hasard de perte et de gain ; et ce que vous jouez est si peu de chose, et de si peu de durée, qu'il y a de la folie à le ménager en cette occasion.

Car il ne sert de rien de dire qu'il est incertain si on gagnera, et qu'il est certain qu'on hasarde ; et que l'infinie distance qui est entre la certitude de ce qu'on expose et l'incertitude de ce que l'on gagnera égale le bien fini qu'on expose certainement à l'infini qui est incertain. Cela n'est pas ainsi : tout joueur hasarde avec certitude pour gagner avec incertitude ; et néanmoins il hasarde certainement le fini pour gagner incertainement le fini, sans pécher contre la raison. Il n'y a pas infinité de distance entre cette certitude de ce qu'on expose, et l'incertitude du gain ; cela est faux. Il y a à la vérité infinité entre la certitude de gagner et la certitude de perdre. Mais l'incertitude de gagner est proportionnée à la certitude de ce qu'on hasarde selon la proportion des hasards de gain et de perte : et de là vient que s'il y a autant de hasards d'un côté que de l'autre, le parti est à jouer égal contre égal ; et alors la certitude de ce qu'on expose est égale à l'incertitude de ce qu'on expose est égale à l'incertitude du gain, tant s'en faut qu'elle en soit infiniment distante. Et ainsi notre proposition est dans une force infinie, quand il n'y a que le fini à hasarder à un jeu où il y a pareils hasards de gain que de perte, et l'infini à gagner. Cela est démonstratif, et si les hommes sont capables de quelques vérités ils le doivent être de celle là.

Je le confesse, je l'avoue. Mais encore n'y aurait-il point de moyen de vois un peu plus clair ? Oui, par le moyen de l'Écriture, et par toutes les autres preuves de la Religion qui sont infinies.

Ceux qui espèrent leur salut, direz vous, sont heureux en cela. Mais ils ont pour contrepoids la crainte de l'enfer.

Mais qui a plus sujet de craindre l'enfer, ou celui qui est dans l'ignorance s'il y a un enfer, et dans la certitude la damnation s'il y en a ; ou celui qui est dans une certaine persuasion qu'il y a un enfer, et dans l'espérance d'être sauvé s'il est ?

Quiconque n'ayant plus que huit jours à vivre ne jugerait pas que le parti de croire que tout cela n'est pas un coup de hasard, aurait entièrement perdu l'esprit. Or si les passions ne nous tenaient point, huit jours et cent ans sont une même chose.

Quel mal vous arrivera-t-il en prenant ce parti ? Vous serez fidèle, honnête, humble, reconnaissant, bienfaisant, sincère, véritable. A la vérité vous ne serez point dans les plaisirs empestés, dans la gloire, dans les délices. Mais n'en aurez vous point d'autre ? Je vous dis que vous y gagnerez en cette vie ; et qu'à chaque pas que vous ferez dans ce chemin, vous verrez tant de certitude du gain, et tant de néant dans ce que vous hasarderez, que vous connaîtrez à la fin que vous avez parié pour une chose certaine et infinie, et que vous n'avez rien donné pour l'obtenir.

L'argument du pari peut être décrit plus précisément de la manière suivante :

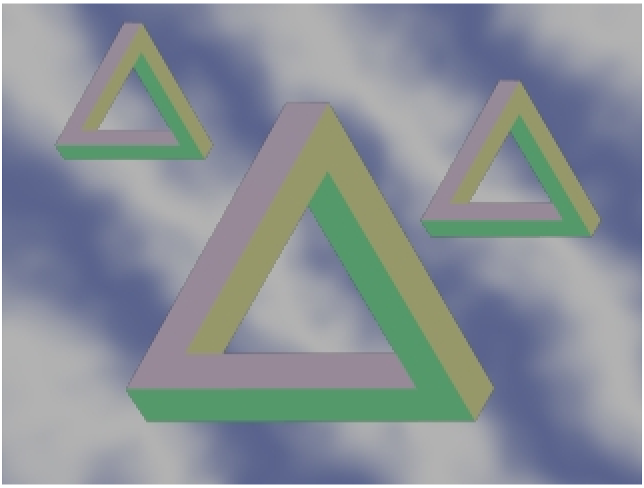
| | | |
|---|---|---|
| (1) | soit Dieu existe soit Dieu n'existe pas | *dichotomie 1* |
| (2) | je peux parier soit pour l'existence de Dieu soit pour sa non-existence | *dichotomie 2* |
| (3) | si je parie en faveur de l'existence de Dieu et que Dieu existe | *cas 1* |
| (4) | alors j'obtiens un gain infini | *de (3)* |
| (5) | si je parie en faveur de l'existence de Dieu et que Dieu n'existe pas | *cas 2* |
| (6) | alors il s'ensuit une perte nulle | *de (5)* |
| (7) | si je parie en faveur de la non-existence de Dieu et que Dieu existe | *cas 3* |
| (8) | alors il en résulte une perte infinie | *de (7)* |
| (9) | si je parie en faveur de la non- | *cas 4* |

| | | |
|---|---|---|
| | existence de Dieu et que Dieu n'existe pas | |
| (10) | alors il ne s'ensuit ni gain ni perte | *de (9)* |
| (11) | il est rationnel d'effectuer un choix afin de maximiser les gain et les pertes attendues | *prémisse* |
| (12) | si je parie en faveur de l'existence de Dieu | *de (3),(5)* |
| (13) | alors le gain maximal est infini et la perte maximale est nulle | *de (4),(6)* |
| (14) | si je parie en faveur de la non-existence de Dieu | *de (7),(9)* |
| (15) | alors le gain maximal est nul et la perte maximale est infinie | *de (8),(10)* |
| (16) | ∴ il est rationnel de parier en faveur de l'existence de Dieu | *de (11), (13),(15)* |

L'argument du pari de Pascal a donné lieu à un certain nombre d'objections. Certains critiques, tels que Jeffrey dans son ouvrage *The Logic of Decision* publié en 1983, ou bien McClennen dans un essai paru en 1994, ont ainsi mis en cause les étapes (3)-(4) et fait valoir que l'utilité infinie qui résulte du gain attendu en cas de pari en faveur de l'existence de Dieu, ne constitue pas un gain réaliste et ne possède donc pas un réel intérêt pratique.

En outre, d'autres auteurs ont souligné que l'attitude intrinsèque qui est sous-tendue par le pari est elle-même critiquable. Voltaire en particulier a jugé cette attitude inconvenante, car elle consiste à décider d'un sujet aussi grave que l'existence de Dieu, exclusivement en fonction de considérations d'intérêt. Dans la situation qui est celle du pari, Voltaire considère ainsi que l'on possède bien les éléments rationnels pour décider de l'existence de Dieu, mais qu'on ne dispose toutefois pas des éléments moraux.

## 25. L'argument selon le Mal



L'argument selon le mal (*argument from evil*) est un argument qui tend à démontrer la non-existence de Dieu. Sa formulation en est très simple. L'argument selon le mal repose sur le fait que le mal est présent dans le monde. La présence de la souffrance et de la douleur constituent une des caractéristiques de notre monde actuel. Pire encore, des atrocités, des crimes horribles se produisent tous les jours dans le monde. L'argument selon le mal considère ces faits indéniables et conclut que cela démontre que Dieu n'existe pas. Il existe plusieurs formulations de l'argument selon le mal. Selon l'une d'elles, Dieu, par définition est un être parfait. Dieu, en outre, est le créateur de toutes choses. Pourtant le mal évident qui existe dans le monde constitue l'une de ces choses. Par conséquent, selon cette variation de l'argument, Dieu est le créateur du mal. Si tel est le cas, l'affirmation selon laquelle Dieu est parfait se trouve ainsi contredite. Cette dernière contradiction entraîne la conclusion que Dieu n'existe pas. Les différentes étapes de l'argument selon le mal peuvent être ainsi décrites de la manière suivante :

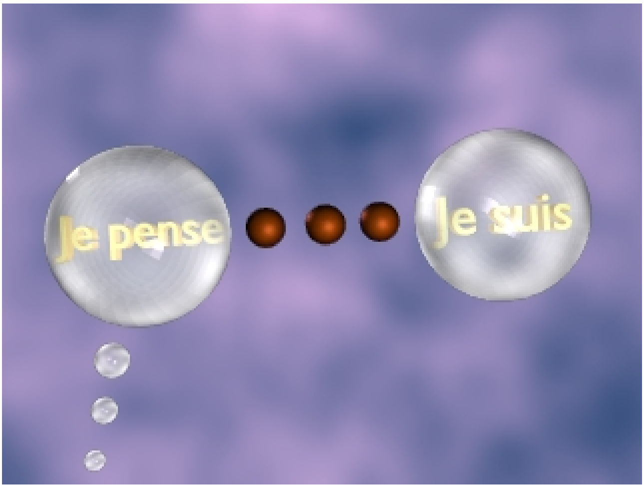
| | | |
|---|---|---|
| (1) | Dieu est parfait | *définition* |
| (2) | Dieu est le créateur de tout ce qui existe | *définition* |
| (3) | le mal existe dans le monde | *prémisse* |
| (4) | Dieu est le créateur du mal qui existe dans le monde | *de (2),(3)* |
| (5) | Dieu n'est pas parfait | *de (4)* |
| (6) | ∴ Dieu n'existe pas | *de (1),(5)* |

Une autre formulation de l'argument selon le mal met l'accent sur la toute-puissance de Dieu, et en particulier sur la notion d'*omnipotence*. L'argument considère que si Dieu existe, alors Dieu est tout-puissant et possède notamment le pouvoir de faire disparaître le mal. Pourtant il s'avère que le mal existe dans le monde, en contradiction avec l'hypothèse selon laquelle Dieu existe. Il en résulte ainsi la conclusion que Dieu n'existe pas. Cette variante de l'argument selon le mal peut ainsi décrite :

| | | |
|---|---|---|
| (7) | si Dieu existe | *hypothèse* |
| (8) | alors Dieu est tout puissant | *définition* |
| (9) | alors Dieu a le pouvoir de supprimer le mal | *de (8)* |
| (10) | si Dieu existe alors Dieu a le pouvoir de supprimer le mal | *de (7),(9)* |
| (11) | le mal existe dans le monde | *prémisse* |
| (12) | Dieu n'a pas le pouvoir de supprimer le mal | *de (10),(11)* |
| (13) | ∴ Dieu n'existe pas | *de (10),(12)* |

L'argument selon le mal a fait l'objet d'objections à la fois anciennes et récentes. Selon une objection récente, soulevée par Alvin Plantinga dans son ouvrage *God and Other Minds* publié en 1967, l'argument n'est pas valide car il repose sur la prémisse fausse selon laquelle Dieu crée le mal ou bien possède le pouvoir de supprimer le mal. Plantinga considère à l'inverse que le libre-arbitre est une vertu nécessaire et que par conséquent, Dieu vise à permettre le développement du libre-arbitre chez les humains. Selon Plantinga, Dieu n'est donc pas responsable du mal (en le créant ou le rendant possible) car le mal résulte directement de l'exercice de choix humains. Et ces choix – qu'ils soient bons ou mauvais – effectués par les hommes sont eux-mêmes indispensables au développement du libre-arbitre.

# 26. Le cogito cartésien




L'*argument du cogito* est dû à Descartes. Il peut être formulé de manière à la fois très brève et très simple : « Je pense, donc je suis ». Cependant, afin d'appréhender exactement la portée du cogito cartésien, il est nécessaire d'en étudier davantage la structure et le contexte.

La formulation originale du cogito se trouve dans le *Discours de la méthode* (Quatrième partie) :

> Je ne sais si je dois vous entretenir des premières méditations que j'y ai faites ; car elles sont si métaphysiques et si peu communes, qu'elles ne seront peut-être pas au goût de tout le monde : et toutefois, afin qu'on puisse juger si les fondements que j'ai pris sont assez fermes, je me trouve en quelque façon contraint d'en parler. J'avais dès longtemps remarqué que pour les mœurs il est besoin quelquefois de suivre des opinions qu'on sait être fort incertaines, tout de même que si elles étaient indubitables, ainsi qu'il a été dit ci-dessus : mais pour ce qu'alors je désirais vaquer seulement à la recherche de la vérité, je pensai qu'il fallait que je fisse

> tout le contraire, et que je rejetasse comme absolument faux tout ce en quoi je pourrais imaginer le moindre doute, afin de voir s'il ne resterait point après cela quelque chose en ma créance qui fut entièrement indubitable. Ainsi, à cause que nos sens nous trompent quelquefois, je voulus supposer qu'il n'y avait aucune chose qui fût telle qu'ils nous la font imaginer ; et parce qu'il y a des hommes qui se méprennent en raisonnant, même touchant les plus simples matières de géométrie, et y font des paralogismes, jugeant que j'étais sujet a faillir autant qu'aucun autre, je rejetai comme fausses toutes les raisons que j'avais prises auparavant pour démonstrations ; et enfin, considérant que toutes les mêmes pensées que nous avons étant éveillés nous peuvent aussi venir quand nous dormons, sans qu'il y en ait aucune pour lors qui soit vraie, je me résolus de feindre que toutes les choses qui m'étaient jamais entrées en l'esprit n'étaient non plus vraies que les illusions de mes songes. Mais aussitôt après je pris garde que, pendant que je voulais ainsi penser que tout était faux, il fallait nécessairement que moi qui le pensais fusse quelque chose ; et remarquant que cette vérité, *je pense, donc je suis*, était si ferme et si assurée, que toutes les plus extravagantes suppositions des sceptiques n'étaient pas capables de l'ébranler, je jugeai que je pouvais la recevoir sans scrupule pour le premier principe de la philosophie que je cherchais.

Il est tentant, à ce stade, de considérer que l'argument du cogito peut être formulé très brièvement : « Je pense, donc je suis » et que sa structure peut être ainsi décrite :

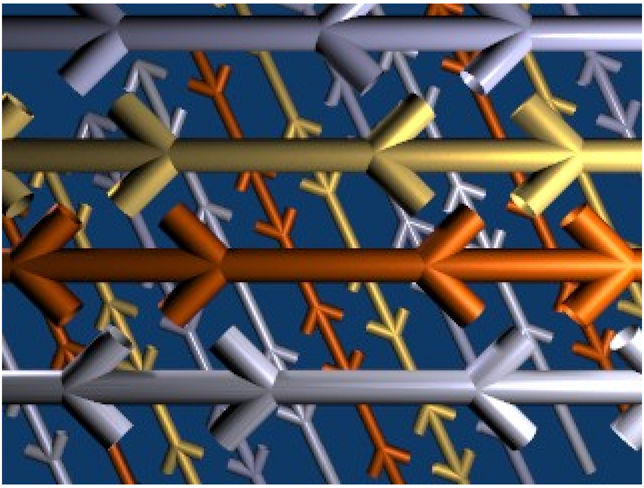
(1) je pense
(2) si je pense alors j'existe
(3) ∴ j'existe

Cependant, il s'agit là d'une interprétation de l'argument de Descartes qui se révèle restrictive. Il apparaît en effet préférable de décrire le cogito cartésien d'une manière qui en capture mieux l'essence, en prenant davantage en compte le contexte de doute dans lequel intervient l'argument lui-même. Car le cogito constitue un argument qui tend à démontrer l'existence de soi, en prenant en compte la possibilité d'être soi-même trompé sur ses pensées ou ses propres perceptions. Descartes va jusqu'à envisager l'hypothèse où l'objet de ses propres pensées est faux, c'est-à-dire où il est trompé sur l'existence même des choses sensibles qui l'entourent, par exemple parce qu'il rêve. Mais même dans cette hypothèse, la conclusion qu'il existe s'impose également à Descartes. La force de l'argument réside ainsi dans le fait que même si j'admets que je suis actuellement trompé par mes propres pensées parce que leur objet est faux, il s'ensuit que j'existe par le fait même que mes pensées sont erronées. Par conséquent, ce que démontre finalement l'argument du cogito, c'est que je ne peux être trompé sur le fait même que j'existe, que mes pensées soient trompeuses ou non. Ainsi, l'argument du cogito peut-il être restitué plus précisément de la manière suivante :

| | | |
|---|---|---|
| (4) | l'objet de mes pensées est soit vrai soit faux | *dichotomie* |
| (5) | si l'objet de mes pensées est vrai | *hypothèse 1* |
| (6) | alors je pense | *conséquence 1* |
| (7) | si l'objet de mes pensées est faux | *hypothèse 2* |
| (8) | alors je pense | *conséquence 2* |
| (9) | je pense | *de (4),(6),(8)* |
| (10) | si je pense alors j'existe | *de (9)* |
| (11) | ∴ j'existe | *de (9),(10)* |

L'argument du cogito constitue une des applications du doute méthodologique mis en œuvre par Descartes. Ce dernier entreprend ainsi de douter de la réalité de toutes les connaissances qu'il a acquises antérieurement et qu'il a toujours tenues pour certaines, non pas parce qu'il remet véritablement en question leur existence, mais parce qu'une telle méthode lui permet de parvenir, de manière optimale, à des connaissances tout à fait certaines et mieux assurées. L'argument du cogito constitue ainsi une illustration de ce doute méthodologique, qui permet à Descartes, dans un tel contexte, d'obtenir une connaissance ferme et assurée, qui correspond à la certitude de sa propre existence.

# 27. L'argument de Lewis Caroll



L'argument de Lewis Caroll a été publié en 1895 dans la revue *Mind*. L'argument y est présenté sous la forme d'un dialogue entre Achille et la tortue. Le problème qui résulte de cet argument peut être présenté de la façon suivante. On considère les étapes du raisonnement suivantes :

| | | |
|---|---|---|
| (1) | deux choses qui sont égales à une troisième sont elles-mêmes égales | *prémisse* |
| (2) | les côtés AB et AC d'un triangle ABC sont tous deux égaux à la longueur DE | *prémisse* |
| (Z) | ∴ les côtés AB et AC du triangle ABC sont égaux | *de (1),(2)* |

A ce stade, un tel raisonnement apparaît tout à fait valide. Mais considérons maintenant l'argument suivant, qui comporte une étape (3) supplémentaire :

| | | |
|---|---|---|
| (1) | deux choses qui sont égales à une | *prémisse* |

| | | |
|---|---|---|
| | troisième sont elles-mêmes égales | |
| (2) | les côtés AB et AC d'un triangle ABC sont tous deux égaux à la longueur DE | *prémisse* |
| (3) | si (1) et (2) sont vraies alors (Z) est vraie | *de (1),(2)* |
| (Z) | ∴ les côtés AB et AC du triangle ABC sont égaux | *de (1),(2), (3)* |

Avant d'affirmer la conclusion (Z), ne convient-il pas préalablement de reconnaître l'étape (3) comme vraie ? L'étape (3) considère que le raisonnement qui conduit à (Z) est valide. Il s'agit là d'une étape nécessaire pour établir que (Z) est vraie. Car si l'étape (3) se révélait fausse, on ne pourrait pas légitimement conclure que (Z) est vraie. Par conséquent, il apparaît légitime de replacer cette étape dans le raisonnement qui conduit à (Z). A ce stade toutefois, il apparaît que si l'on rétablit l'étape (3), on se doit également de prendre en compte nouvelle étape supplémentaire (4), qui conduit à considérer l'ensemble du raisonnement suivant :

| | | |
|---|---|---|
| (1) | deux choses qui sont égales à une troisième sont elles-mêmes égales | *prémisse* |
| (2) | les côtés AB et AC d'un triangle ABC sont tous deux égaux à la longueur DE | *prémisse* |
| (3) | si (1) et (2) sont vraies alors (Z) est vraie | *de (1),(2)* |
| (4) | si (1), (2) et (3) sont vraies alors (Z) est vraie | *de (1),(2), (3)* |
| (Z) | ∴ les côtés AB et AC du triangle ABC sont égaux | *de (1),(2), (3),(4)* |

Mais à nouveau, il apparaît que le raisonnement précédent peut être prolongé, en incorporant une nouvelle étape supplémentaire :

| | | |
|---|---|---|
| (1) | deux choses qui sont égales à une troisième sont elles-mêmes égales | *prémisse* |
| (2) | les côtés AB et AC d'un triangle ABC sont tous deux égaux à la longueur DE | *prémisse* |
| (3) | si (1) et (2) sont vraies alors (Z) est vraie | *de (1),(2)* |
| (4) | si (1), (2) et (3) sont vraies alors (Z) est vraie | *de (1),(2), (3)* |
| (5) | si (1), (2), (3) et (4) sont vraies alors (Z) est vraie | *de (1),(2), (3),(4)* |
| (Z) | ∴ les côtés AB et AC du triangle ABC sont égaux | *de (1),(2), (3),(4),(5)* |

Un tel raisonnement peut être prolongé *ad infinitum* et il en résulte ainsi une régression infinie. Par conséquent, il s'ensuit que l'on ne parvient jamais à la conclusion (Z).

L'argument de Lewis Caroll repose sur le fait qu'avant de parvenir à la conclusion (Z), il convient d'admettre que le raisonnement qui conduit à cette conclusion est valide. De manière générale, l'argument – à l'instar du *paradoxe de la course* de Zénon d'Elée, mais aussi du *paradoxe d'Achille et la tortue*, qui est un autre paradoxe de Zénon – souligne qu'avant de parvenir à la conclusion (Z), on doit parcourir une série infinie d'étapes et que dans ces conditions, on ne parvient jamais à formuler la conclusion (Z).

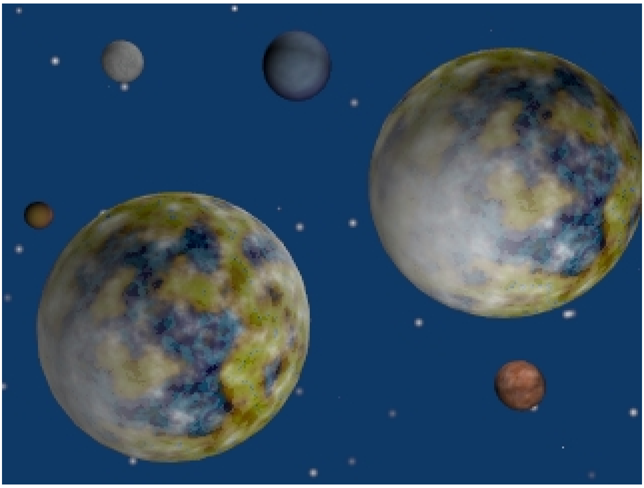
L'argument de Lewis Caroll souligne l'importance du *modus ponens*. Cette règle d'inférence autorise le raisonnement dont la structure est la suivante (P et Q étant deux propositions) :

| | | |
|---|---|---|
| (6) | P est vraie | *prémisse* |
| (7) | si P est vraie alors Q est vraie | *prémisse* |
| (8) | ∴ Q est vraie | *de (6),(7)* |

L'argument souligne ainsi le fait qu'avant d'appliquer une règle d'inférence telle que le *modus ponens*, il est nécessaire de disposer d'une seconde règle décrivant comment on doit appliquer le modus ponens, puis d'une troisième règle décrivant comment on doit appliquer la règle qui décrit comment appliquer le modus ponens, et ainsi de suite. Une régression infinie s'ensuit.

Une objection qui a été opposée classiquement à l'argument de Caroll est qu'un tel problème ne survient pas dans la logique formelle, où chaque règle se trouve formalisée. Dans ce cas, le mécanisme déductif se réduit alors à une manipulation de symboles. Toutefois, un tel système formaliste présente l'inconvénient de ne pas prendre en compte l'aspect sémantique des choses, pourtant essentiel. Car ce dernier aspect se révèle totalement absent de ce qui ne se réduit alors qu'à une manipulation de caractères symboliques dépourvus de sens.

## 28. L'expérience de pensée de la Terre jumelle



L'expérience de pensée de la Terre jumelle a été introduite par Hilary Putnam, dans un essai publié en 1975. Putnam y expose trois expériences de pensée, et l'une d'entre elles – l'expérience de pensée $H_2O$-XYZ – introduit le problème de la Terre jumelle. Putnam y met en scène une planète, la Terre jumelle, qui se révèle en tous points identique à la Terre, à une seule différence près. Cette différence concerne le corps composé qui est dénommé « eau » sur la Terre et dont la structure atomique est $H_2O$. Sur la Terre jumelle, il existe en effet un corps composé qui possède toutes les propriétés de notre *eau*, telles que le fait d'être liquide, transparent, inodore, etc. mais dont la composition chimique est XYZ. Appelons *eau** un tel corps composé. Sur la Terre jumelle, les habitants appellent également « eau » ce dernier corps composé. A ce stade, selon Putnam, il apparaît que *eau* se réfère au corps composé $H_2O$ et que *eau** se réfère au corps composé XYZ. Ainsi, *eau* et *eau** sont respectivement utilisés d'une manière tout à fait identique par les terriens et les habitants de la Terre jumelle. De plus, le contenu des pensées d'un habitant de la Terre ou bien de la Terre jumelle est tout à fait identique lorsqu'ils pensent respectivement à l'*eau* ou bien à l'*eau**. Par conséquent, il s'ensuit que le contenu sémantique de leurs pensées respectives ne peut être déterminé d'une manière purement interne, et ne peut donc être élucidé qu'en recourant à une donnée *externe*. C'est ici que se situe, selon Putnam, la leçon du problème posé par l'expérience de la Terre jumelle. On peut en effet s'interroger pour savoir si la signification, le contenu sémantique d'un mot ou d'un concept se trouve ou non exclusivement dans notre cerveau. Selon Putnam, ce que démontre l'expérience de la Terre jumelle, c'est qu'une réponse négative doit être apportée à cette question. Car seul le recours à une donnée externe permet dans l'expérience de la Terre jumelle de déterminer le contenu sémantique des pensées d'un Terrien et d'un habitant de la Terre jumelle lorsqu'ils pensent ou ils parlent respectivement de l'*eau* ou bien de l'*eau**. Ainsi, conclut Putnam, il convient d'adopter une conception *externaliste* pour la détermination du contenu mental.

Le raisonnement auquel conduit l'expérience de la Terre jumelle peut être ainsi détaillé :

| | | |
|---|---|---|
| (1) | sur Terre, il existe un corps composé liquide, transparent, inodore, etc. dont la composition est $H_2O$ | *prémisse* |
| (2) | le corps composé dont la composition est $H_2O$ est l'*eau* | *définition* |
| (3) | sur la Terre jumelle, il existe un corps composé liquide, transparent, inodore, etc. dont la composition est XYZ | *prémisse* |
| (4) | le corps composé dont la composition est XYZ est l'*eau** | *définition* |
| (5) | les habitants de la Terre appellent « eau » le corps composé dont la | *prémisse* |

| | | |
|---|---|---|
| | composition est $H_2O$ | |
| (6) | les habitants de la Terre jumelle appellent « eau » le corps composé dont la composition est XYZ | *prémisse* |
| (7) | le contenu des pensées d'un habitant de la Terre lorsqu'il pense à l'*eau* est *x* | *de (1),(5)* |
| (8) | le contenu des pensées d'un habitant de la Terre jumelle lorsqu'il pense à l'*eau** est *x* | *de (2),(6)* |
| (9) | le contenu des pensées d'un habitant de la Terre ou de la Terre jumelle lorsqu'ils pensent respectivement à l'*eau* ou à l'*eau** est identique | *de (7),(8)* |
| (10) | ∴ il faut recourir à une donnée externe pour différencier le contenu sémantique des pensées d'un Terrien qui pense à l'*eau* de celui des pensées d'un habitant de la Terre jumelle qui pense à l'*eau** | *de (9)* |

A ce stade, il apparaît que la portée du problème soulevé par Putnam s'étend au-delà de la seule expérience de pensée de la Terre jumelle et de notre concept d'*eau*. Car un raisonnement analogue peut s'appliquer à toutes les catégories d'objets désignés par notre langage usuel : un nuage, une montagne, une chaise, etc. Pour chacun de nos objets usuels et familiers, il s'avère ainsi nécessaire, en vertu de l'expérience de la Terre jumelle, de recourir à un critère externe afin d'appréhender le contenu sémantique correspondant.

On a pu objecter à l'expérience de pensée de la Terre jumelle que la situation correspondante n'est pas réaliste. En effet, si un corps composé devait posséder des propriétés tout à fait identiques à celles de notre *eau*, sa composition ne devrait-elle pas alors être la même que celle de l'*eau*, c'est-à-dire $H_2O$. Selon cette objection, la proposition (3) selon laquelle il existe sur une autre planète un corps composé possédant des propriétés identiques à celles de l'*eau* et dont la composition chimique est différente, se révèle irréaliste, voire contradictoire.

## 29. L'argument contre le principe de vérifiabilité

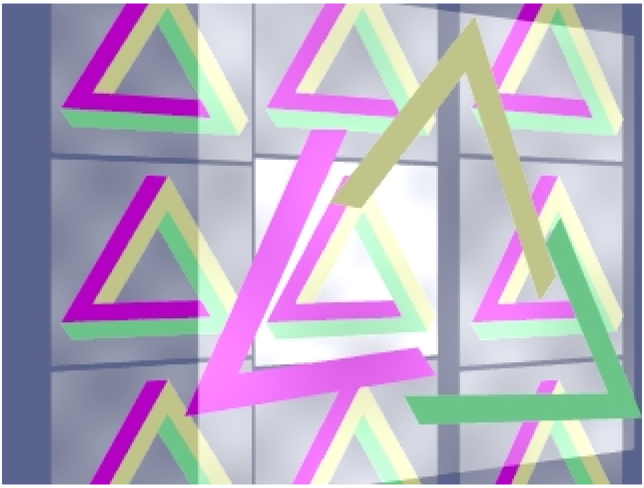
L'argument basé sur le principe de vérifiabilité résulte des travaux d'un groupe de philosophes appartenant au courant de pensée du positivisme logique. Ce courant de pensée s'inscrit dans le cadre des idées émises dans les années 1920-1930 par le cercle de Vienne, qui comprenait notamment Rudolf Carnap et Kurt Gödel. Le positivisme logique distingue deux types de propositions porteuses de sens : certaines propositions (a) sont analytiques, alors que d'autres (b) peuvent être vérifiées de manière expérimentale. Les propositions analytiques sont par exemple les propositions mathématiques, telles que « un chien est un mammifère » (qui est analytiquement vraie) ou bien « un triangle possède deux angles droits » (qui est analytiquement fausse), dont on peut établir la véracité ou la fausseté par la seule déduction. A l'inverse, les propositions vérifiables expérimentalement peuvent être confirmées ou infirmées de manière empirique. Ainsi, « je mesure 1,73 mètre » ou bien « Proxima du Centaure se situe à 4,23 années-lumière de la Terre » constituent des propositions qui peuvent être vérifiées de manière expérimentale. Tout autre type de proposition, c'est-à-dire qui n'est ni analytique ni vérifiable expérimentalement, est dépourvue de sens. Le positivisme logique, influencé par les idées développées par Ludwig Wittgenstein, conduit ainsi au rejet des propositions métaphysiques, considérées comme non significatives, car elles ne satisfont pas à l'un des deux critères précédents. Selon ce point de vue, les affirmations métaphysiques ne possèdent pas de fondement logique, car elles ne satisfont pas le critère de *vérifiabilité*, en vertu duquel toute affirmation doit pouvoir être vérifiée expérimentalement. De ce point de vue, une affirmation métaphysique devrait pourvoir fait l'objet d'une confirmation ou d'une infirmation. Tel n'est cependant pas le cas et par conséquent, les affirmations métaphysiques doivent être rejetées.

Cependant, un tel argument basé sur le principe de vérifiabilité a fait l'objet de l'objection suivante, due notamment à Ewing, dans son ouvrage *The fundamental questions of philosophy* paru en 1962 : le principe de vérifiabilité lui-même n'est pas vérifiable expérimentalement. Ainsi, le principe de vérifiabilité ne satisfait pas lui-même au critère de vérifiabilité. Car on ne dispose pas d'un procédé permettant de vérifier ce dernier de manière expérimentale. Ainsi le principe de vérifiabilité se trouve-t-il victime du principe-même qu'il prétend édicter. Ceci montre comment un tel principe se révèle en fait trop restrictif. L'argument contre le principe de vérifiabilité peut être ainsi décrit étape par étape de la manière suivante :

| | | |
|---|---|---|
| (1) | soit le principe de vérifiabilité prévaut, soit il ne prévaut pas | *dichotomie* |
| (2) | en vertu du principe de vérifiabilité, toute affirmation doit être vérifiable | *prémisse* |
| (3) | être vérifiable, pour une proposition, | *définition* |

| | | |
|---|---|---|
| | consiste dans le fait qu'il est possible de la confirmer ou de l'infirmer | |
| (4) | le principe de vérifiabilité ne peut être confirmé expérimentalement | *prémisse* |
| (5) | le principe de vérifiabilité ne peut être infirmé expérimentalement | *prémisse* |
| (6) | le principe de vérifiabilité ne peut être ni confirmé ni infirmé expérimentalement | *de (4),(5)* |
| (7) | le principe de vérifiabilité n'est pas vérifiable | *de (3),(6)* |
| (8) | ∴ le principe de vérifiabilité ne prévaut pas | *de (7)* |

## 30. L'allégorie de la caverne



La célèbre *allégorie de la caverne* a été décrite par Platon dans la République (Livre VII). Platon y met en scène des humains qui ont été enchaînés, depuis leur enfance, aux murs d'une caverne. Ces prisonniers sont enchaînés d'une manière telle qu'ils ne peuvent pas bouger la tête et ne peuvent donc pas se voir les uns les autres. Cependant, la caverne communique par une ouverture avec l'extérieur. Tout ce que peuvent observer ces prisonniers ne sont que des reflets de personnes et d'animaux qui passent à l'extérieur de la caverne, ainsi que les ombres de fleurs, de rochers, etc. tels qu'ils apparaissent sur les murs de la caverne. Pour les prisonniers, la réalité se limite aux ombres et aux reflets qu'ils observent sur ces murs. Mais un jour, un des prisonniers parvient à briser ses chaînes et à s'échapper de la caverne. Il sort alors pour la première fois de la caverne, et à la lumière du jour, découvre alors les véritables personnes, les animaux réels, les fleurs authentiques, etc., dans leurs formes et leurs couleurs originales. Il n'a plus alors qu'une seule idée : retourner dans la caverne et informer ses anciens compagnons que ce qu'ils voient sur les murs de la caverne ne sont que des reflets, des ombres et des apparences d'un autre niveau de réalité qui leur apparaîtrait s'ils brisaient eux aussi leurs liens et s'en allaient à la lumière du jour. Retournant dans la caverne, il entreprend d'expliquer à ses compagnons enchaînés que ce qu'ils voient n'est que le reflet de la réalité véritable. Mais ses anciens compagnons ne le croient pas, et finissent par le tuer. L'allégorie a clairement la structure d'une analogie. Car pour Platon, les ombres qui apparaissent sur les murs de la caverne représentent le monde des apparences. A l'inverse, les objets véritables tels qu'on peut les observer à la lumière du jour appartiennent au monde des Idées.

L'extrait de la République qui comprend l'allégorie de la caverne met en scène le dialogue suivant entre Socrate et Glaucon :

> SOCRATE - Maintenant, représente-toi notre nature selon qu'elle a été instruite ou ne l'a pas été, sous des traits de ce genre : imagine des hommes dans une demeure souterraine, une caverne, avec une large entrée, ouverte dans toute sa longueur à la lumière : ils sont là les jambes et le cou enchaînés depuis leur enfance, de sorte qu'ils sont immobiles et ne regardent que ce qui est devant eux, leur chaîne les empêchant de tourner la tête. La lumière leur parvient d'un feu qui, loin sur une hauteur, brûle derrière eux ; et entre le feu et les prisonniers s'élève un chemin en travers duquel imagine qu'un petit mur a été dressé, semblable aux cloisons que des montreurs de marionnettes placent devant le public, au-dessus desquelles ils font voir leurs marionnettes.
> GLAUCON - Je vois.
> - Imagine le long du mur des hommes qui portent toutes sortes d'objets qui dépassent le mur ; des statuettes d'hommes et d'animaux, en pierre, en bois, faits de toutes sortes de matériaux ; parmi ces porteurs, naturellement il y en a qui parlent et d'autres qui se taisent.
> - Voilà un étrange tableau et d'étranges prisonniers.

- Ils nous ressemblent. Penses-tu que de tels hommes aient vu d'eux-mêmes et des uns et des autres autre chose que les ombres projetées par le feu sur la paroi de la caverne qui leur fait face ?
- Comment cela se pourrait-il, en effet, s'ils sont forcés de tenir la tête immobile pendant toute leur vie ?
- Et pour les objets qui sont portés le long du mur, est-ce qu'il n'en sera pas de même ?
- Bien sûr.
- Mais, dans ces conditions, s'ils pouvaient se parler les uns aux autres, ne penses-tu pas qu'ils croiraient nommer les objets réels eux-mêmes en nommant ce qu'ils voient ?
- Nécessairement.
- Et s'il y avait aussi dans la prison un écho que leur renverrait la paroi qui leur fait face ? Chaque fois que l'un de ceux qui se trouvent derrière le mur parlerait, croiraient-ils entendre une autre voix, à ton avis, que celle de l'ombre qui passe devant eux ?
- Ma foi non.
- Non, de tels hommes ne penseraient absolument pas que la véritable réalité puisse être autre chose que les ombres des objets fabriqués.
- De toute nécessité.
- Envisage maintenant ce qu'ils ressentiraient à être délivrés de leurs chaînes et à être guéris de leur ignorance, si cela leur arrivait, tout naturellement, comme suit : si l'un d'eux était délivré et forcé soudain de se lever, de tourner le cou, de marcher et de regarder la lumière ; s'il souffrait de faire tous ces mouvements et que, tout ébloui, il fût incapable de regarder les objets dont il voyait auparavant les ombres, que penses-tu qu'il répondrait si on lui disait que jusqu'alors il n'a vu que des futilités mais que, maintenant, plus près de la réalité et tourné vers des êtres plus réels, il voit plus juste ; lorsque, enfin, en lui montrant chacun des objets qui passent, on l'obligerait à force de questions à dire ce que c'est, ne penses-tu pas qu'il serait embarrassé et trouverait que ce qu'il voyait auparavant était plus véritable que ce qu'on lui montre maintenant ?
- Beaucoup plus véritable.

- Si on le forçait à regarder la lumière elle-même, ne penses-tu pas qu'il aurait mal aux yeux, qu'il la fuirait pour se retourner vers les choses qu'il peut voir et les trouverait vraiment plus distinctes que celles qu'on lui montre ?
- Si.
- Mais si on le traînait de force tout au long de la montée rude, escarpée, et qu'on ne le lâchât pas avant de l'avoir tiré dehors à la lumière du soleil, ne penses-tu pas qu'il souffrirait et s'indignerait d'être ainsi traîné ; et que, une fois parvenu à la lumière du jour, les yeux pleins de son éclat, il ne pourrait pas discerner un seul des êtres appelés maintenant véritables ?
- Non, du moins pas sur le champ.

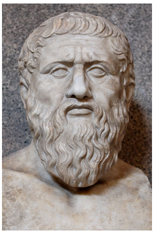
- Il aurait, je pense, besoin de s'habituer pour être en mesure de voir le monde d'en haut. Ce qu'il regarderait le plus facilement d'abord, ce sont les ombres, puis les reflets des hommes et des autres êtres sur l'eau, et enfin les êtres eux-mêmes. Ensuite il contemplerait plus facilement pendant la nuit les objets célestes et le ciel lui-même – en levant les yeux vers la lumière des étoiles et de la lune – qu'il ne contemplerait, de jour, le soleil et la lumière du soleil.
- Certainement.
- Finalement, je pense, c'est le soleil, et non pas son image dans les eaux ou ailleurs, mais le soleil lui-même à sa vraie place, qu'il pourrait voir et contempler tel qu'il est.
- Nécessairement.
- Après cela il en arriverait à cette réflexion, au sujet du soleil, que c'est lui qui produit les saisons et les années, qu'il gouverne tout dans le monde visible, et qu'il est la cause, d'une certaine manière, de tout ce que lui-même et les autres voyaient dans la caverne.
- Après cela, il est évident que c'est à cette conclusion qu'il en viendrait.

- Mais quoi, se souvenant de son ancienne demeure, de la science qui y est en honneur, de ses compagnons de captivité, ne penses-tu pas qu'il serait heureux de son changement et qu'il plaindrait les autres ?
- Certainement.
- Et les honneurs et les louanges qu'on pouvait s'y décerner mutuellement, et les récompenses qu'on accordait à qui distinguait avec le plus de précision les ombres qui se présentaient, à qui se rappelait le mieux celles qui avaient l'habitude de passer les premières, les dernières, ou ensemble, et à qui était le plus capable, à partir de ces observations, de présager ce qui devait arriver : crois-tu qu'il les envierait ? Crois-tu qu'il serait jaloux de ceux qui ont acquis honneur et puissance auprès des autres, et ne préférerait-il pas de loin endurer ce que dit Homère : « être un valet de ferme au service d'un paysan pauvre », plutôt que de partager les opinions de là-bas et de vivre comme on y vivait.
- Oui, je pense qu'il accepterait de tout endurer plutôt que de vivre comme il vivait.
- Et réfléchis à ceci : si un tel homme redescend et se rassied à la même place, est-ce qu'il n'aurait pas les yeux offusqués par l'obscurité en venant brusquement du soleil ?
- Si, tout à fait.
- Et s'il lui fallait à nouveau donner son jugement sur les ombres et rivaliser avec ces hommes qui ont toujours été enchaînés, au moment où sa vue est trouble avant que ses yeux soient remis – cette réaccoutumance exigeant un certain délai – ne prêterait-il pas à rire, ne dirait-on pas à son propos que pour être monté là-haut, il en est revenu les yeux gâtés et qu'il ne vaut même pas la peine d'essayer d'y monter ; et celui qui s'aviserait de les délier et de les emmener là-haut, celui-là s'ils pouvaient s'en emparer et le tuer, ne le tueraient-ils pas ?
- Certainement.

On peut détailler, à ce stade, les différentes étapes qui sous-tendent l'allégorie de la caverne :

| | | |
|---|---|---|
| (1) | les prisonniers de la caverne sont convaincus que les objets qu'ils observent quotidiennement sous les objets réels | *prémisse* |
| (2) | les prisonniers de la caverne observent en réalité sur les murs les ombres et les reflets des objets véritables | *de (1)* |
| (3) | la situation des prisonniers de la caverne est analogue à notre situation présente | *analogie* |
| (4) | nous sommes convaincus que les objets que nous observons quotidiennement sous les objets réels | *prémisse* |
| (5) | ∴ les objets que nous observons ne sont en réalité que les ombres et les reflets des objets véritables | *de (2),(3), (4)* |

La conclusion de Platon est que la situation humaine est analogue à celle des prisonniers de la caverne. En ce sens, l'allégorie de la caverne est clairement un argument par analogie. Cependant, à ce stade, la conclusion qui en



résulte peut être diversement interprétée. On peut distinguer ainsi deux interprétations principales. Selon la première interprétation, les prisonniers de la caverne sont les hommes, et les objets que voient ceux-ci ne sont que le pâle reflet des objets authentiques, qui sont les Idées ou Archétypes. Il existe ainsi des Archétypes du nombre « 7 », du courage et de la tolérance, d'un lion et du soleil, etc. dans le monde des Idées. En ce sens, les humains croient que la réalité ultime est celle qui correspond à leurs perceptions, alors que ceci est illusoire et que la réalité véritable se situe au niveau des Archétypes. Ainsi, nous évoluons tous les jours dans ce qui ne constitue que le second plan correspondant à la projection des objets authentiques, eux-mêmes situés au premier plan, c'est-à-dire celui des Archétypes. En ce sens, l'allégorie de la caverne se révèle proche de l'expérience des cerveaux dans une cuve et de son illustration moderne à travers le film Matrix.

Un second type d'interprétation peut toutefois être appliqué à l'allégorie de la caverne. Une telle interprétation est directement liée à la théorie de la connaissance de Platon. Car Platon distingue la connaissance née de l'opinion et la connaissance authentique. Ainsi, les connaissances des êtres et des objets que possèdent les prisonniers de la caverne ne sont que des connaissances tirées de l'opinion. Il ne s'agit pas de connaissances véritables, car elles sont façonnées, transformées et déformées par l'éducation qui a été reçue par chacun. Les connaissances usuelles que nous possédons sont, selon Platon, perverties par le tumulte des passions humaines, l'ambition, la compétition, les idées reçues, etc. À l'inverse, les connaissances authentiques et véritables se situent au-delà des passions, des haines, des honneurs, des idées établies. Selon Platon, chaque humain doit s'élever ainsi au-dessus des passions qui l'enchaînent, afin de parvenir à la connaissance véritable.

# 31. L'argument de la simulation

L'*argument de la simulation* (*simulation argument*) a été décrit très récemment par Nick Bostrom, dans un article publié en 2003 dans la revue *Philosophical Quarterly*. L'argument repose essentiellement sur le fait qu'il apparaît assez probable qu'une civilisation post-humaine procédera à des simulations d'humains. Il apparaît vraisemblable en effet que des civilisations post-humaines très avancées, disposeront à la fois des capacités et de la volonté de réaliser des simulations d'humains extrêmement réalistes. Si tel était le cas, le nombre des humains simulés devrait alors excéder très largement le nombre des humains authentiques. Dans un tel cas, il s'ensuit que la prise en compte du fait que chacun de nous existe conduit à considérer comme plus probable que nous appartenons aux humains simulés qu'aux humains authentiques. Selon Bostrom, la conclusion qui résulte de l'argument de la simulation est que la probabilité de chacune des trois assertions suivantes est d'environ 1/3 :

(1) l'humanité est vouée à une extinction prochaine
(2) une civilisation post-humaine ne réalisera pas de simulations d'humains
(3) nous vivons actuellement dans une simulation

Ces probabilités ne sont pas étonnantes en ce qui concerne les assertions (1) et (2), mais la probabilité relative à l'assertion (3) en vertu de laquelle nous vivons actuellement dans une simulation, se révèle tout à fait contraire à l'intuition.

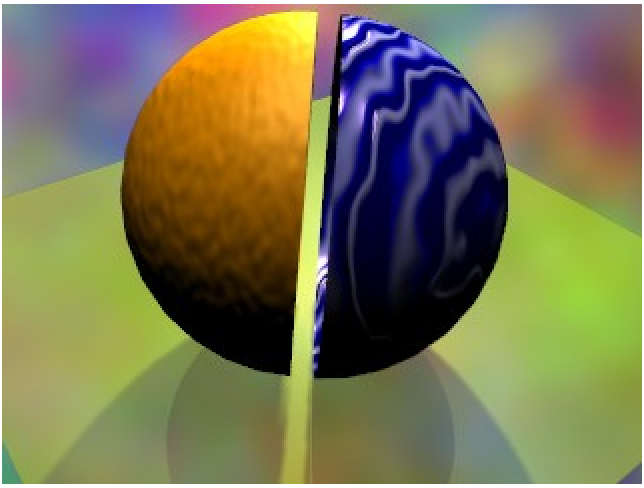
L'argument de la simulation est également exposé de manière plus succincte par Brian Weatherson, dans une réponse à l'article original de Bostrom, publiée en 2004. Selon ce dernier, le noyau véritable de l'argument de la simulation peut être décrit de la façon suivante. Tout d'abord, il est très probable qu'une civilisation post-humaine sera apte à produire des simulations réalistes d'êtres humains. De même, il est très probable que le nombre des êtres humains simulés excédera largement le nombre des humains réels. Ainsi, à un âge post-humain, le ratio entre les *humains simulés* et les *humains véritables* devrait être largement en faveur des *humains simulés*. A ce stade, il apparaît que le simple fait de prendre en compte notre existence actuelle conduit à considérer qu'il est probable que nous soyons des *humains simulés*. Ceci invite à penser que la probabilité que nos pensées, nos impressions, nos sensations, etc. soient le résultat d'une simulation, est élevée.

La conclusion de l'argument de la simulation, d'une manière tout à fait similaire à l'argument de l'Apocalypse, se révèle contraire à l'intuition et au bon sens. Cependant, de la même manière que pour l'argument de l'Apocalypse, la tâche qui consiste à déterminer avec précision l'étape fallacieuse au niveau de l'argument de la simulation, se révèle très difficile.

Une première objection qui pourrait être soulevée à l'encontre de l'argument de la simulation porte sur la nécessite de faire appel à un principe d'indifférence (en vertu duquel il n'y a lieu a priori de privilégier ici aucune des hypothèses). Car l'humain que nous sommes est-il véritablement choisi de manière aléatoire au sein de la classe de référence qui inclut à la fois les humains et les humains simulés. Il semble en effet que l'argument de la simulation ne vaille que si nous sommes choisis de manière aléatoire au sein de la classe de référence. N'y a-t-il pas là le même problème que celui qui apparaît en présence de l'argument de l'Apocalypse ? Bostrom, cependant, répond à cette objection en faisant valoir que le principe d'indifférence utilisé dans l'argument de la simulation n'est pas de même nature que celui auquel se réfère l'argument de l'Apocalypse. En effet, dans l'argument de l'Apocalypse, une prémisse importante est que chaque humain, compte tenu de son rang de naissance, doit être considéré comme choisi de manière aléatoire au sein de la classe de référence. Dans l'argument de la simulation, le principe d'indifférence utilisé se révèle plus faible, car il est appliqué sans aucune considération de rang de naissance (ou de tout autre critère de même nature), mais procède à partir de la simple constatation de notre existence en tant que membres de la classe de référence.

Une autre objection qui peut être soulevée est que l'argument de la simulation est lui-même auto-réfutant. En effet, si sa conclusion est vraie, il s'ensuit que l'argument lui-même est le produit d'une simulation et que l'ensemble de notre logique est elle-même simulée. Dans ce cas, on ne peut donc retenir comme valables les conclusions qui résultent de l'argument. Toutefois, on peut remarquer qu'une telle objection vaut aussi pour l'argument du rêve, l'expérience des cerveaux dans une cuve, etc. Ainsi, une telle objection apparaît-elle trop générale, et il semble qu'elle ne réponde pas, de manière précise, au problème spécifique posé par l'argument de la simulation.

## 32. L'argument dualiste en vertu de la divisibilité



Dans le cours des *Méditations métaphysiques* (*Sixième méditation*), Descartes développe un argument qui se propose de prouver l'existence de la dualité corps/esprit. Il se propose ainsi de montrer comment le corps et l'esprit constituent deux composantes essentielles de l'homme, dont la nature s'avère cependant fondamentalement différente. Cet argument prend place dans le débat qui oppose le *matérialisme* à l'*idéalisme*. Le matérialisme est la doctrine selon laquelle seules les choses matérielles et physiques existent. Dans ce cadre, les phénomènes de nature mentale se réduisent uniquement à des phénomènes d'origine matérielle. Ainsi, selon le matérialisme, tout ce qui existe est matière et peut être caractérisé en termes purement physiques. A l'opposé, l'idéalisme est le point de vue selon lequel seules les choses de nature mentale existent. Dans ce contexte, les choses matérielles ne possèdent d'existence qu'à travers nos propres perceptions. Selon le point de vue idéaliste, tout ce qui existe se réduit ainsi à une existence purement mentale. Le matérialisme et l'idéalisme constituent des points de vue monistes. A l'inverse, le *dualisme* constitue un point de vue pluraliste qui considère que les choses de nature physique et mentale existent à la fois. Selon ce point de vue, le mental et le physique, dont la nature profonde est fondamentalement différente, coexistent. Le point de vue dualiste a été défendu de manière célèbre par Descartes. Car il existe, selon Descartes, une dualité corps/esprit, qui constitue la contrepartie applicable à l'homme du dualisme physique/mental. Descartes fonde son argumentation sur les propriétés respectives du corps et de l'esprit, qui sont fondamentalement distinctes. Il considère ainsi que la matière physique qui constitue notre corps possède une extension dans l'espace et se révèle par conséquent divisible. A l'inverse, l'esprit, selon Descartes, ne possède pas d'extension spatiale et ne présente donc pas cette même propriété de divisibilité. Ainsi, le corps et l'esprit présentent au moins une propriété différente et sont donc, en vertu de la loi de Leibnitz – selon laquelle deux objets sont identiques si et seulement si toutes leurs propriétés sont identiques – fondamentalement distincts.

L'argument dualiste en vertu de la divisibilité provient du passage suivant des *Méditations métaphysiques* :

> Pour commencer donc cet examen, je remarque ici, premièrement, qu'il y a une grande différence entre l'esprit et le corps, en ce que le corps, de sa nature, est toujours divisible, et que l'esprit est entièrement indivisible. Car en effet, lorsque je considère mon esprit, c'est-à-dire moi-même en tant que je suis seulement une chose qui pense, je n'y puis distinguer aucunes parties, mais je me conçois comme une chose seule et entière. Et quoique tout l'esprit semble être uni à tout le corps, toutefois un pied, ou un bras, ou quelque autre partie étant séparée de mon corps, il est certain que pour cela il n'y aura rien de retranché de mon esprit. Et les facultés de vouloir, de sentir, de concevoir, etc., ne peuvent pas

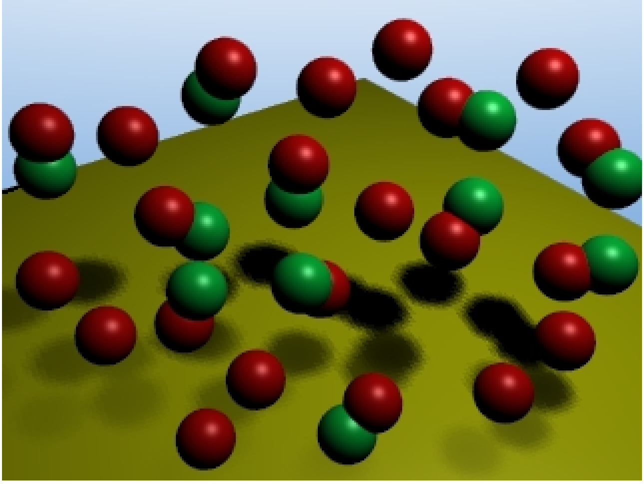
> proprement être dites ses parties : car le même esprit s'emploie tout entier à vouloir, et aussi tout entier à sentir, à concevoir, etc. Mais c'est tout le contraire dans les choses corporelles ou étendues : car il n'y en a pas une que je ne mette aisément en pièces par ma pensée, que mon esprit ne divise fort facilement en plusieurs parties et par conséquent que je ne connaisse être divisible. Ce qui suffirait pour m'enseigner que l'esprit ou l'âme de l'homme est entièrement différente du corps, si je ne l'avais déjà d'ailleurs assez appris.

Les différentes étapes de l'argument dualiste de Descartes en vertu de la divisibilité peuvent être détaillées de la façon suivante :

| | | |
|---|---|---|
| (1) | mon corps possède une extension dans l'espace | *prémisse* |
| (2) | tout ce qui possède une extension dans l'espace est divisible | *prémisse* |
| (3) | mon corps est divisible | *de (1),(2)* |
| (4) | mon esprit ne possède pas d'extension dans l'espace | *prémisse* |
| (5) | mon esprit n'est pas divisible | *prémisse* |
| (6) | mon corps et mon esprit possèdent au moins une propriété différente | *de (3),(5)* |
| (7) | deux choses sont identiques si et seulement si elles possèdent des propriétés identiques | *loi de Leibnitz* |
| (8) | si deux choses possèdent des propriétés différentes alors ces deux choses sont distinctes | *de (7)* |
| (9) | ∴ mon corps et mon esprit sont deux choses distinctes | *de (6),(8)* |

Le point de vue dualiste de Descartes a donné lieu à une objection importante qui est la suivante : s'il existe une dualité corps/esprit, comment ces deux composantes fondamentalement différentes d'un même être humain interagissent-elles ? La nature de l'interaction qui résulte de la doctrine de la dualité corps/esprit, n'a jusqu'à présent pas été élucidée. Il s'agit là d'une lacune importante dans la doctrine dualiste, car une théorie dualiste complète se doit de décrire de manière explicite les modalités de l'interaction entre le corps et l'esprit.

## 33. Le problème de la Belle au bois dormant



Le problème de la Belle au bois dormant (*Sleeping Beauty Problem*) a suscité un certain nombre de discussions récentes, en particulier entre Adam Elga et David Lewis dans des articles respectivement publiés en 2000 et en 2001 dans la revue *Analysis*. Le problème de la Belle au bois dormant a été ainsi décrit de la manière suivante par Elga. Des chercheurs ont préparé une expérience pendant laquelle ils se proposent d'endormir la Belle au bois dormant. Celle-ci sera endormie deux jours durant : lundi et mardi. Toutefois, pendant son sommeil, elle sera réveillée soit une fois, soit deux fois. Le nombre de fois où elle sera réveillée dépendra du résultat du lancer d'une pièce de monnaie parfaitement équilibrée. Si la pièce tombe sur face, la Belle ne sera réveillée qu'une fois, le lundi. En revanche, si la pièce tombe sur pile, elle sera réveillée deux fois, lundi et mardi. Dans les deux cas, après avoir été réveillée le lundi, la Belle sera à nouveau endormie et elle oubliera qu'elle a été réveillée. Compte tenu de ces éléments, lorsque la Belle est réveillée, à quel degré doit-elle croire que la pièce est tombée sur face ?

A ce stade, il apparaît qu'un premier type (I) de raisonnement conduit à penser que la probabilité que la pièce soit tombée sur face est égale à 1/2. En effet, la pièce de monnaie est équilibrée, et par conséquent, si l'expérience est répétée, il en résultera un nombre à peu près égal de lancers face ou de lancers pile. La probabilité initiale de face ou de pile est donc 1/2. Mais lorsque la Belle est réveillée, elle ne reçoit aucune information nouvelle. Par conséquent, elle n'a aucune raison de modifier sa croyance initiale. Car il aurait été rationnel de modifier des probabilités initiales si des données nouvelles lui avaient été fournies. Mais tel n'est pas le cas et par conséquent, la Belle ne possède aucune justification pour modifier ses probabilités initiales. Un tel raisonnement correspond, de manière simplifiée, à celui qui est mis en œuvre par David Lewis.

Il s'avère cependant qu'un second type (II) de réponse apparaît possible. Le raisonnement correspondant conduit à la conclusion que la probabilité que la pièce soit tombée sur face est 1/3. Il faut imaginer que l'expérience est répétée de nombreuses fois. Dans ce cas, il s'avérera qu'environ 1/3 des réveils seront des réveils qui se produiront alors que la pièce est tombée sur face. Et de même, environ 2/3 des réveils se produiront alors que la pièce est tombée sur pile. Ainsi, lorsque la Belle est réveillée, elle peut considérer valablement qu'il s'agit d'un réveil consécutif à un lancer face avec une probabilité de 1/3. Par conséquent, la Belle doit conclure que la probabilité que la pièce de monnaie est tombée sur face est de 1/3.

Il est utile de formaliser les éléments du problème de la Belle au bois dormant, de manière à en mettre en évidence la structure interne. Le problème est en effet basé sur les deux hypothèses concurrentes suivantes :

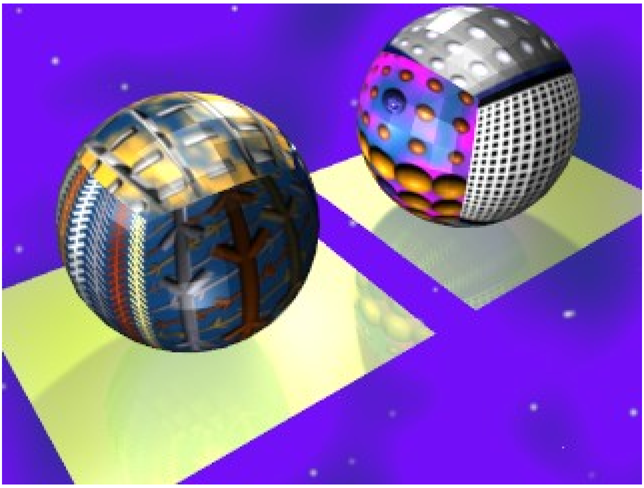
(H1) la Belle sera réveillée une seule fois (FACE)
(H2) la Belle sera réveillée deux fois (PILE)

De même, il apparaît que trois cas sont possibles :

(a) la pièce est tombée sur FACE et la Belle est réveillée le lundi
(b) la pièce est tombée sur PILE et la Belle est réveillée le lundi
(c) la pièce est tombée sur PILE et la Belle est réveillée le mardi

Le problème qui résulte de la situation correspondant au problème de la Belle au bois dormant est que les deux raisonnements (I) et (II) paraissent a priori valides, alors qu'ils conduisent à des conclusions contradictoires. Ainsi, l'un des deux raisonnements doit être fallacieux. Mais lequel ? Et pourquoi ? Dans la littérature contemporaine relative au problème de la Belle au bois dormant, les deux raisonnements concurrents possèdent leurs défenseurs et leurs détracteurs, et il n'existe pas actuellement de solution consensuelle.

## 34. L'argument du mauvais génie



L'*argument du mauvais génie* est un argument célèbre décrit par Descartes dans les *Méditations métaphysiques*. L'argument du mauvais génie constitue un argument en faveur du scepticisme. L'argument proprement dit repose sur une expérience de pensée. Descartes envisage ainsi l'hypothèse selon laquelle un mauvais génie existe, qui est capable de le tromper non seulement au niveau de toutes ses perceptions sensorielles, mais également au niveau de l'ensemble de ses connaissances, y compris celles qui concernent les mathématiques. Considérant qu'il ne possède pas la certitude absolue qu'un tel mauvais génie n'existe pas, Descartes conclut qu'il est donc possible que toutes ses connaissances soient fausses et qu'il est fondé à douter de la sorte de l'ensemble de ces dernières.

L'argument du mauvais génie apparaît dans le passage suivant des *Méditations métaphysiques* (*Première méditation*) :

> Je supposerai donc qu'il y a, non point un vrai Dieu, qui est la souveraine source de vérité, mais un certain mauvais génie, non moins rusé et trompeur que puissant qui a employé toute son industrie à me tromper. Je penserai que le ciel, l'air, la terre, les couleurs, les figures, les sons et toutes les choses extérieures que nous voyons, ne sont que des illusions et tromperies, dont il se sert pour surprendre ma crédulité. Je me considérerai moi-même comme n'ayant point de mains, point d'yeux, point de chair, point de sang, comme n'ayant aucuns sens, mais croyant faussement avoir toutes ces choses. Je demeurerai obstinément attaché à cette pensée ; et si, par ce moyen, il n'est pas en mon pouvoir de parvenir à la connaissance d'aucune vérité, à tout le moins il est en ma puissance de suspendre mon jugement. C'est pourquoi je prendrai garde soigneusement de ne point recevoir en ma croyance aucune fausseté, et préparerai si bien mon esprit à toutes les ruses de ce grand trompeur, que, pour puissant et rusé qu'il soit, il ne pourra jamais rien imposer.

L'argument du mauvais génie peut être détaillé de la manière suivante :

| | | |
|---|---|---|
| (1) | il est possible qu'il existe un mauvais génie, capable de me tromper sur l'ensemble de mes perceptions sensorielles et de mes connaissances mathématiques | *hypothèse* |
| (2) | si je suis trompé au niveau de l'ensemble de mes perceptions sensorielles et de mes connaissances mathématiques (par exemple le fait que je me trouve actuellement devant le feu de la cheminée ou que la somme des angles d'un triangle est égale à un angle plat) alors l'ensemble de mes croyances sont fausses | *prémisse* |
| (3) | il est possible que l'ensemble de mes | *de (1),(2)* |

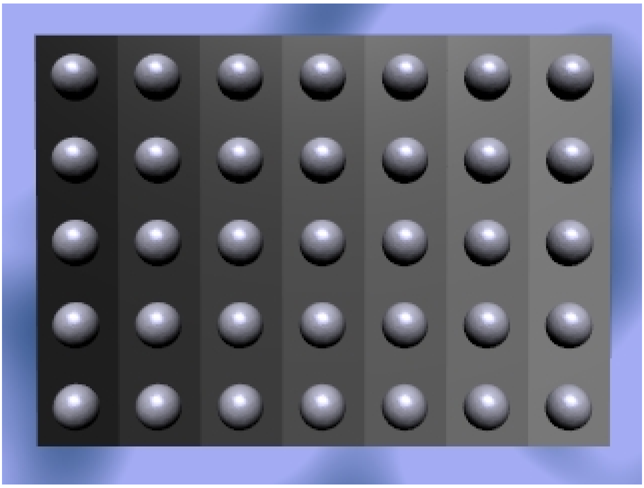
| | | |
|---|---|---|
| | croyances soient fausses | |
| (4) | si je ne possède pas la certitude qu'un tel mauvais génie n'existe pas, alors je ne peux pas considérer que l'ensemble de mes croyances sont vraies | *de (1),(3)* |
| (5) | je ne possède pas la certitude qu'un tel mauvais génie n'existe pas | *prémisse* |
| (6) | je ne peux pas considérer que l'ensemble de mes croyances sont vraies | *de (4),(5)* |
| (7) | ∴ je suis fondé à douter de l'ensemble de mes croyances | *de (6)* |

L'argument vise clairement les connaissances *a posteriori* qui s'appliquent aux objets matériels (par exemple une table, un cheval ou la planète Saturne), mais également les connaissances *a priori* telles que celles qui résultent des mathématiques (par exemple le fait que la somme des angles d'un triangle est égale à un angle plat, ou bien 1 + 3 = 4).

On peut douter cependant que l'argument du mauvais génie autorise un doute généralisé, c'est-à-dire qu'il s'applique à l'ensemble de nos connaissances. En effet, ainsi que le démontre Descartes lui-même, il semble qu'une proposition telle que « je pense, donc je suis » échappe à un tel doute à portée universelle. En ce sens, la conclusion de l'argument du mauvais génie se révèle trop forte. Toutefois, il apparaît que même si on restreint ainsi la portée de la conclusion de l'argument, l'essentiel de celui-ci demeure et permet encore de conclure en faveur du scepticisme.

Une autre objection qui peut être formulée par rapport à l'argument du mauvais génie est que l'argument est auto-réfutant. Car celui-ci s'applique à la fois aux connaissances a posteriori et a priori. Or la conclusion qui résulte de l'argument du mauvais génie lui-même constitue une connaissance a priori. Je suis donc autorisé à douter également de cette dernière conclusion. Ainsi, l'argument lui-même est-il ébranlé par sa propre conclusion. Étant donné que je suis fondé à douter de l'ensemble de mes connaissances a priori, je suis ainsi fondé à douter que je peux douter de l'ensemble de mes croyances.

## 35. L'argument de la chambre chinoise de Searle



L'argument de la chambre chinoise a été décrit par John Searle, dans un article paru en 1980 dans la revue *Behavioral and Brain Sciences*. Cet argument repose sur une expérience de pensée, qui est la suivante. Supposez que vous n'avez aucune connaissance de la langue chinoise et vous vous trouvez enfermé, seul, dans une chambre qui ne contient que les objets suivants : (a) un jeu de textes dactylographiés en langue chinoise, intitulé « le script » ; (b) un second jeu de documents en langue chinoise, intitulé « l'histoire », accompagnés d'une série de règles en français permettant de mettre en relation les premiers documents avec les seconds ; (c) un troisième jeu de documents, intitulé « les questions », comportant des symboles en chinois ainsi que des instructions en français permettant de mettre en relation les symboles chinois avec les deux premiers jeux de documents. A ce moment précis, un texte en chinois vous est transmis sous la porte.

Consultant alors vos quatre jeux de documents, vous rédigez alors un autre texte en langue chinoise, intitulé « les réponses », que vous transmettez à votre tour sous la porte de la chambre.

L'expérience de pensée de Searle est basée sur une analogie. Elle met en parallèle la situation qui est celle de la personne qui se trouve dans la chambre, avec la situation correspondant à un programme d'ordinateur effectuant une traduction. La personne qui se trouve dans la chambre reçoit un texte rédigé en chinois, puis, consultant une série de documents, rédige à son tour un nouveau document en langue chinoise, qui constitue une réponse au premier document reçu. Une telle réponse n'est pas différente de celle qu'aurait faite une personne ayant une excellente compréhension de la langue chinoise. Et ceci souligne combien la véritable compréhension du texte en chinois qui lui a été soumis lui échappe en fait complètement. Car la personne qui se trouve dans la chambre est capable de répondre de manière compétente à la question qui lui est posée, mais ignore totalement le contenu de cette réponse. L'expérience a ainsi pour but de mettre en évidence comment le contenu *sémantique* du texte échappe à la machine, alors même qu'elle possède la maîtrise de son contenu *syntaxique*.

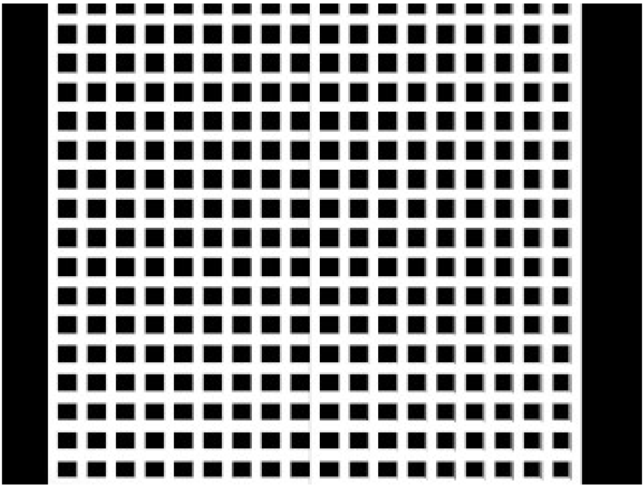
L'argument de Searle a pour but de constituer une objection au point de vue selon lequel un programme d'ordinateur est capable de penser. Ce dernier point de vue constitue la thèse dite de « l'IA forte » (intelligence artificielle forte). Selon cette dernière thèse, les ordinateurs possèdent réellement une aptitude à penser, de la même manière que le font les humains. En ce sens, un programme d'ordinateur peut posséder une véritable compréhension d'une situation donnée. L'IA forte s'oppose ainsi à la thèse de l'IA faible, en vertu de laquelle les programmes d'ordinateur ne constituent que des simulations de l'esprit humain. En ce sens, le résultat d'un programme d'ordinateur ne constitue pas un authentique processus de pensée, mais une simple simulation, aussi réussie soit-elle, de ce dernier.

L'argument de Searle proprement dit, illustré par l'expérience de la chambre chinoise, peut être ainsi détaillé :

| | | |
|---|---|---|
| (1) | soit l'IA forte prévaut, soit l'IA faible prévaut | *dichotomie* |
| (2) | les programmes d'ordinateur font usage de symboles | *prémisse* |
| (3) | les symboles correspondent au contenu syntaxique d'un texte | *prémisse* |
| (4) | l'esprit humain fait usage du contenu sémantique d'un texte | *prémisse* |
| (5) | l'expérience de la chambre chinoise montre que le contenu syntaxique d'un texte ne suffit pas à déterminer le contenu sémantique d'un texte | *de (3),(4)* |
| (6) | la situation de la personne dans la chambre chinoise est analogue à celle d'un programme d'ordinateur effectuant une traduction | *analogie* |
| (7) | ∴ les programmes d'ordinateur ne parviennent pas à déterminer le contenu sémantique d'un texte | *de (5),(6)* |
| (8) | l'IA forte ne prévaut pas | *de (7)* |
| (9) | ∴ c'est l'IA faible qui prévaut | *de (1),(8)* |

L'argument de la pièce chinoise a engendré une énorme controverse. Bien que Searle réponde par avance dans son article original à un certain nombre d'objections, son argument n'a pas convaincu de nombreux auteurs. Toutefois, aucun d'entre eux n'est parvenu à indiquer, d'une manière qui se révèle consensuelle, l'étape précise dans l'argumentation de Searle qui se révèle défectueuse.

# 36. Le test de Turing



Alan Turing, dans un article célèbre paru en 1950 dans la revue *Mind*, se propose d'élucider la question : « Les machines peuvent-elles penser ? ». Au lieu d'essayer de répondre à cela de manière classique en définissant les notions de « machine » et de « penser », Turing s'oriente vers une autre voie. Il s'attache ainsi à décrire le jeu suivant, qu'il appelle le jeu de l'imitation :

> *Le jeu de l'imitation* Ce jeu se joue à trois personnes : un homme (A), une femme (B) et un interrogateur (C) de l'un ou l'autre sexe. L'interrogateur se trouve dans une pièce différente de celle où se trouvent les deux autres. Le but du jeu pour l'interrogateur est de parvenir à déterminer quelle personne parmi les deux autres personnes est l'homme ou la femme. L'interrogateur connaît chacune d'entre elles par la dénomination X et Y et à la fin du jeu, il doit dire soit « X est A et Y est B », soit « X est B et Y est A ». Dans ce but, l'interrogateur est autorisé à poser des questions à A et à B.

De nos jours, la version originale du *jeu de l'imitation* décrite par Turing est habituellement remplacée par une expérience simplifiée qui est la suivante :

> *Le jeu de l'imitation (version moderne)* Ce jeu se joue à deux personnes et une machine : un homme (A), une machine (M) et un interrogateur (C). A et C sont de l'un ou l'autre sexe. L'interrogateur se trouve connecté à A et à M à l'aide d'un terminal, par l'intermédiaire duquel ils peuvent communiquer. Toutefois, l'interrogateur ne peut voir ni l'homme ni la machine et ne sait donc pas qui est l'humain et qui est la machine. Sa mission est de s'attacher à déterminer qui est l'humain et qui est la machine, en leur posant des questions. L'interrogateur se trouve dans une pièce différente de celle où se trouvent les deux autres. La machine et l'humain cherchent à convaincre l'interrogateur que chacun d'eux est humain. Le but du jeu pour l'interrogateur est de parvenir à déterminer qui est véritablement l'humain. Si l'interrogateur ne parvient pas à distinguer l'humain de la machine, on considère alors que la machine est intelligente.

On peut remarquer qu'une version ancienne du test de Turing peut être attribuée à Descartes dans son *Discours de la méthode*, qui imagine une situation de nature similaire, dans le passage suivant[5] :

> Et je m'étais ici particulièrement arrêté à faire voir que s'il y avait de telles machines qui eussent les organes et la figure extérieure d'un singe ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux ; au lieu que s'il y en avait qui eussent la ressemblance de nos corps, et imitassent autant nos actions que moralement il serait possible, nous aurions toujours deux moyens très certains pour reconnaître

5 De : http://abu.cnam.fr/BIB/auteurs/descartesr.html. Avec quelques adaptations.

> qu'elles ne seraient point pour cela de vrais hommes : dont le premier est que jamais elles ne pourraient user de paroles ni d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées. Car on peut bien concevoir qu'une machine soit tellement faite qu'elle profère des paroles, et même qu'elle en profère quelques unes à propos des actions corporelles qui causeront quelque changement en ses organes, comme, si on la touche en quelque endroit, qu'elle demande ce qu'on lui veut dire ; si en un autre, qu'elle crie qu'on lui fait mal, et choses semblables ; mais non pas qu'elle les arrange diversement pour répondre au sens de tout ce qui se dira en sa présence, ainsi que les hommes les plus hébétés peuvent faire. Et le second est que, bien qu'elles fissent plusieurs choses aussi bien ou peut-être mieux qu'aucun de nous, elles manqueraient infailliblement en quelques autres, par lesquelles on découvrirait qu'elles n'agiraient pas par connaissance, mais seulement par la disposition de leurs organes. Car, au lieu que la raison est un instrument universel qui peut servir en toutes sortes de rencontres, ces organes ont besoin de quelque particulière disposition pour chaque action particulière ; d'où vient qu'il est moralement impossible qu'il y en ait assez de divers en une machine pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir.

En second lieu, se fondant sur le jeu de l'imitation, Turing effectue la prédiction suivante. Il considère que d'ici l'an 2000, il sera tout à fait possible de programmer un ordinateur de manière à ce qu'un interrogateur humain moyen n'ait pas plus de 70/100 de chances au jeu de l'imitation d'identifier correctement l'humain et la machine, après avoir posé une série de questions durant 5 minutes. D'une manière générale, le test de Turing a pour finalité de montrer que le temps n'est plus très loin où il sera impossible de différencier l'homme de la machine. Selon Turing, ceci constitue une démonstration que l'intelligence humaine peut être entièrement simulée par ordinateur.

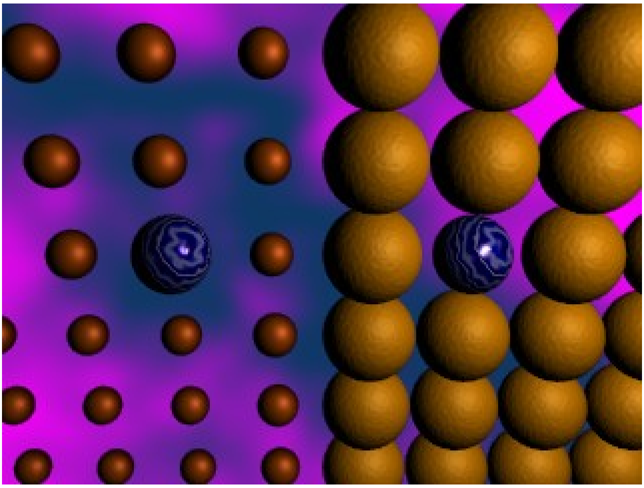
L'argument qui sous-tend le test de Turing peut être présenté ainsi de manière détaillée :

| | | |
|---|---|---|
| (1) | si on effectue un test afin de distinguer l'intelligence humaine de l'intelligence simulée de la machine | *hypothèse* |
| (2) | alors on ne parvient pas à définir un critère permettant d'effectuer une telle distinction | *de (1)* |
| (3) | il est quasiment impossible de discerner l'intelligence humaine de l'intelligence simulée de la machine | *de (2)* |
| (4) | ∴ l'intelligence humaine peut être entièrement simulée | *de (3)* |

Dans ce contexte, l'argument basé sur le test de Turing apparaît étroitement lié à l'argument de la simulation décrit récemment par Nick Bostrom.

On peut objecter à l'expérience de Turing que les potentialités du cerveau humain et de l'intelligence ne commencent qu'à peine à être connues. Ainsi, de nouvelles aptitudes de l'intelligence humaine pourraient bien être découvertes, qui échapperaient alors entièrement au test de Turing. Dans le même ordre d'idées, on peut également considérer que ce que permet de conclure le test de Turing, c'est qu'actuellement et dans un futur proche, il sera assez difficile de discerner une machine d'un être humain. Cependant, cela n'autorise pas à conclure qu'une telle différenciation ne sera jamais possible. Ne s'agit pas là d'une conclusion trop forte ? Pour conclure valablement que l'intelligence humaine peut être entièrement simulée par ordinateur, il faudrait disposer d'une certitude absolue que la différenciation entre l'humain et la machine, dans les conditions du test, ne peut être effectuée.

## 37. Le problème de Gettier



Le *problème de Gettier* a été exposé par Edmund Gettier, dans un article paru en 1963 dans la revue *Analysis*. Classiquement, on considère qu'une personne S *sait* une proposition donnée P dès lors que trois conditions sont simultanément réunies : (a) la proposition P est vraie ; (b) S croit que P est vraie ; (c) S est justifié dans sa croyance que P est vraie. Ainsi, S sait que P s'il possède une croyance vraie et justifiée de P. Cette triple condition du savoir est communément admise. Cependant, Gettier entreprend de montrer que cette triple condition du savoir n'est pas fondée et que ces trois critères ne constituent pas une condition suffisante. Gettier illustre ainsi son propos à l'aide de deux situations concrètes.

Le premier cas concret décri par Gettier est le suivant. Deux personnages, Pierre et Jean, ont tous deux postulé pour un emploi. Pierre possède des éléments décisifs qui l'autorisent à penser que la proposition suivante, dont la structure est celle d'une conjonction, est vraie :

(1) Jean est celui qui obtiendra l'emploi et Jean a dix pièces de monnaie dans sa poche

Les éléments déterminants dont dispose Pierre sont d'une part le fait que le président de la société lui a assuré que ce serait Jean qui aurait l'emploi ; et d'autre part le fait que Pierre a préalablement compté le nombre de pièces – au nombre de dix – qui se trouvaient dans la poche de Jean. Ainsi, (1) a pour conséquence :

(2) celui qui obtiendra l'emploi a dix pièces de monnaie dans sa poche

Dans ce cas, on peut considérer que Pierre sait que (2), puisque la triple condition précitée est satisfaite : la proposition (2) est vraie, Pierre croit que (2) et Pierre se trouve justifié par (1) dans sa croyance que (2). Mais imaginons maintenant que, sans que Pierre le sache, ce soit finalement Pierre lui-même qui ait l'emploi et qu'il possède également dix pièces de monnaie dans sa poche. Dans cette hypothèse, (1) se révèle alors fausse. De plus, il apparaît que Pierre ne sait pas véritablement que (2), alors même que la triple condition du savoir est pourtant satisfaite. Ainsi, il apparaît dans ce cas particulier que Pierre ne sait pas P, bien que les trois conditions précitées soient réunies.

Le second cas pratique de Gettier est le suivant. Soit la proposition suivante :

(3) Jean possède une Ford

De plus, Pierre sait que Jean a toujours possédé une Ford et que ce dernier a récemment effectué un voyage avec lui. Pierre possède ainsi des éléments décisifs en faveur de (3). De plus, il s'avère que Pierre a un autre camarade, Bernard, dont il ignore toutefois un certain nombre de choses. Soient maintenant les trois propositions :

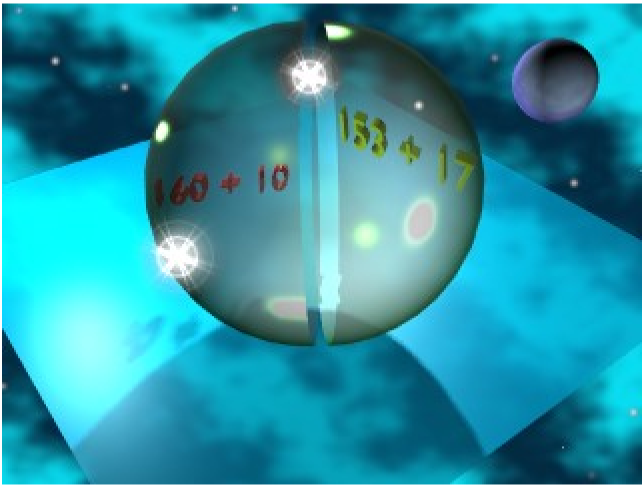
(4) Jean possède une Ford ou Bernard est à Boston
(5) Jean possède une Ford ou Bernard est à Barcelone
(6) Jean possède une Ford ou Bernard est à Paris

A ce stade, il apparaît que chacune de ces trois propositions constitue une conséquence logique de (3). Cependant, on peut considérer que Pierre sait que (4), (5) et (6), car chacune de ces propositions est vraie, et d'autre part, Pierre possède de chacune d'elles une croyance justifiée. Mais maintenant supposons que Bernard ne possède pas de Ford, mais utilise une Chrystler de location et que Bernard se trouve, sans que Pierre ne le sache, à Barcelone. Dans ce cas, il apparaît que Pierre ne sait pas véritablement que (5) est vraie, alors même que la triple condition du savoir concernant (5) se trouve à nouveau satisfaite.

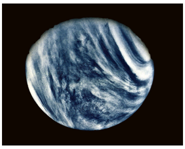
Les deux exemples qui précédent, conclut Gettier, montrent que la triple condition mentionnée plus haut ne constitue pas une condition *suffisante* pour que S sache que P. Cependant, un certain nombre de réponses ont été apportées par rapport au problème de Gettier. L'une de ces réponses souligne que la justification qui est présente dans les deux cas mentionnés par Gettier se révèle insuffisante. Car la connaissance ne doit-elle pas être motivée par une preuve véritable, et non par ce qui ne constitue qu'une justification fragile ? Pierre base en effet sa croyance sur le seul fait que le président de la société lui a assuré que ce serait Jones qui aurait l'emploi. Cependant à ce stade, Pierre possède la certitude des déclarations du président, mais n'a pas la preuve des faits correspondants. Car le président ne pourrait-il pas changer d'avis ultérieurement ? Par conséquent, on peut penser que l'étape de justification se révèle insuffisante. En ce sens, les deux exemples décrits par Gettier se caractérisent par une justification faible, alors que précisément une justification forte s'avère nécessaire. Selon ce type d'objection on le voit, la triple condition de la connaissance demeure acceptable, mais la condition de justification doit être remplacée par une condition plus forte, qui correspond à une *preuve*. Dans ce contexte, la véritable connaissance correspond à une croyance vraie et prouvée. Toutefois, une telle réponse au problème de Gettier ne suffit pas à en dissiper les conséquences. Car ce type de réponse présente l'inconvénient de s'avérer trop radicale. Son application conduit ainsi à ne pas considérer comme conduisant à une connaissance authentique, nombre de situations de la vie courante où l'on ne dispose pas d'une preuve aussi définitive et absolue.

Plusieurs solutions proposées pour résoudre le problème de Gettier ont pour finalité d'empêcher l'émergence des cas décrits par Gettier, en ajoutant une condition supplémentaire. Une des solutions de ce type est basée sur le fait que la connaissance résulte d'une croyance vraie et justifiée, mais aussi que cette triple condition ne peut être obtenue de manière accidentelle. Cette dernière condition a pour but d'empêcher les cas décrits par Gettier de survenir. Mais une telle conception ne s'est pas avérée entièrement satisfaisante, car la définition même des conditions accidentelles est apparue problématique. En effet, dans certains cas, l'apparition accidentelle de la triple condition précitée ne conduit pas à une véritable connaissance, alors que dans d'autres circonstances, la survenue accidentelle de cette triple condition engendre un authentique savoir.

## 38. Le problème de Frege relatif aux propositions d'identité

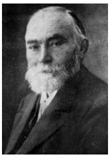



Le problème relatif aux propositions d'identité a été décrit par Gottlob Frege dans son essai *On Sense and Reference* publié en 1892. Ce problème s'établit comme suit. On considère tout d'abord une assertion telle que « l'étoile du matin est l'étoile du soir ». Dans cas, il apparaît que les expressions « l'étoile du matin » et « l'étoile du soir » se réfèrent à un seul et même objet : la planète Vénus. On le voit, la structure de la proposition « l'étoile du matin est l'étoile du soir » présente la forme « A » = « B ». De manière générale, des propositions qui présentent une telle structure sont vraies si et seulement si « A » et « B » se réfèrent à un même objet. Ceci peut également être formulé en termes de nombres. Si l'on considère les expressions « 160 + 10 » et « 153 + 17 », il apparaît que ces deux expressions se réfèrent à un même entier naturel qui est 170. Frege s'est attaché à décrire ainsi une théorie de la vérité pour les



propositions présentant la structure « A » = « B », en définissant les conditions dans lesquelles de telles propositions se révèlent vraies.

Cependant, Frege observa qu'un problème émergeait avec ce type d'analyse. Il est apparu en effet que les conditions dans lesquelles une proposition de la forme « A » = « B » se révélait vraie (les conditions de vérité) étaient identiques à celles dans lesquelles une proposition de la forme « A » = « A » était également vraie. Or une proposition de la forme « A » = « A » telle que « l'étoile du matin est l'étoile du matin » s'avère, d'un point de vue sémantique, très différente d'une proposition telle que « l'étoile du matin est l'étoile du soir ». Il s'ensuit ainsi la conclusion que les conditions de vérité sont identiques, pour des propositions pourtant sémantiquement très différentes de la forme « A » = « B » ou « A » = « A ».

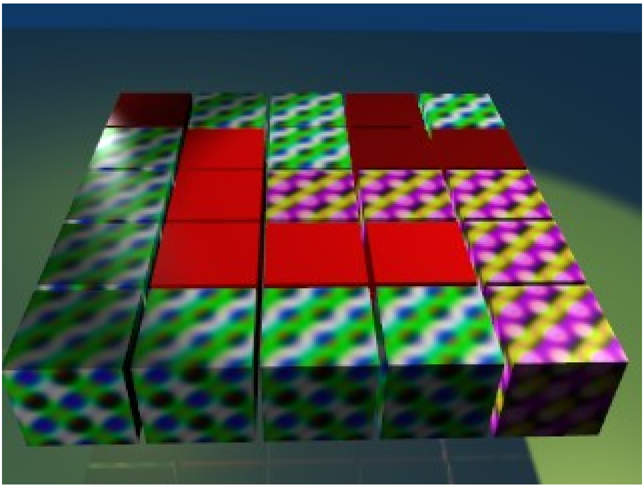


Le raisonnement qui conduit au problème de Frege relatif aux propositions d'identité peut être ainsi formalisé :

| | | |
|---|---|---|
| (1) | l'étoile du matin est l'étoile du soir | *prémisse* |
| (2) | « l'étoile du matin » et « l'étoile du soir » se réfèrent à la planète Vénus | *définition* |
| (3) | (1) est vraie | *de (1),(2)* |
| (4) | l'étoile du matin est l'étoile du matin | *identité* |
| (5) | « l'étoile du matin » se réfère à la planète Vénus | *définition* |
| (6) | (4) est vraie | *de (4),(5)* |
| (7) | (1) présente la structure « A » = « B » | *de (1)* |
| (8) | (4) présente la structure « A » = « A » | *de (4)* |
| (9) | une proposition qui présente la structure « A » = « B » est vraie si et seulement si « A » et « B » se réfèrent à un même objet | *généralisation* |
| (10) | une proposition qui présente la structure « A » = « A » est vraie si et seulement si « A » et « A » se réfèrent | *de (9)* |

| | | |
|---|---|---|
| | à un même objet | |
| (11) | les conditions de vérité d'une proposition qui présente la structure « A » = « B » et d'une proposition qui présente la structure « A » = « A » sont identiques | *de (9),(10)* |
| (12) | d'un point de vue sémantique, une proposition qui présente la structure « A » = « B » est très différente d'une proposition qui présente la structure « A » = « A » | *de (1),(4)* |
| (13) | ∴ les conditions de vérité de deux propositions sémantiquement très différentes sont identiques | *de (11), (12)* |

# 39. Le paradoxe de l'analyse



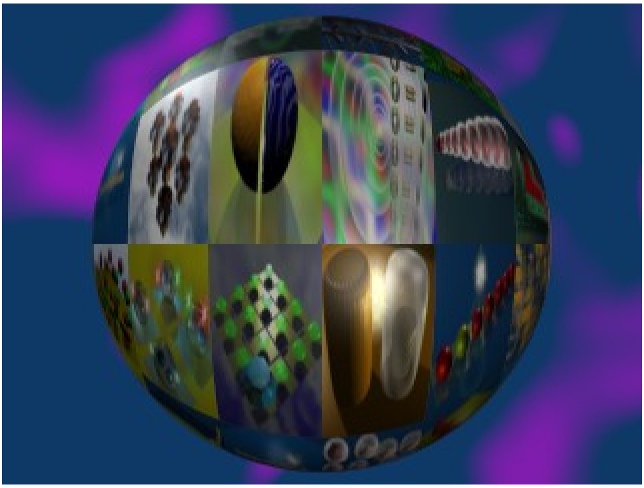
Le *paradoxe de l'analyse* résulte des travaux de George Edward Moore. Le paradoxe est basé sur la démarche méthodologique qui consiste à analyser un concept donné. Appelons α un tel concept. L'analyse de ce concept α présente ainsi la forme : α = E. Ici, α est le concept qui est analysé (l'*analysandum*) alors que E est une expression (l'*analysans*) – plus ou moins complexe – qui définit et décrit le contenu sémantique de α. Le paradoxe émerge dès que l'on considère les deux possibilités qui se présentent : (a) soit l'*analysans* décrit exactement le contenu du concept α ; (b) soit l'*analysans* ne décrit pas exactement le contenu du concept α. Dans la première hypothèse, il s'ensuit que l'analyse effectuée est triviale, et ne présente donc aucun intérêt. Dans la seconde hypothèse, il apparaît que l'*analysans* ne décrit pas exactement le contenu du concept α et par conséquent, l'analyse effectuée est fausse. Ainsi, l'analysans est soit trivial, soit faux. Dans les deux cas, l'analyse effectuée se révèle inutile. Pourtant, ceci est en contradiction avec la donnée qui résulte de notre intuition pré-théorique selon laquelle l'analyse d'un concept donné se révèle le plus souvent utile.

Le raisonnement correspondant au paradoxe de l'analyse peut être ainsi détaillé :

| | | |
|---|---|---|
| (1) | soit l'*analysans* décrit exactement le contenu du concept α, soit l'*analysans* n'en décrit pas exactement le contenu | *dichotomie* |
| (2) | si l'*analysans* décrit exactement le contenu du concept α | *hypothèse 1* |
| (3) | alors l'analyse est triviale | *de (2)* |
| (4) | si l'*analysans* ne décrit pas exactement le contenu du concept α | *hypothèse 2* |
| (5) | alors l'analyse est inexacte | *de (4)* |
| (6) | l'analyse du concept α est soit triviale soit inexacte | *de (3),(5)* |
| (7) | ∴ l'analyse du concept α est inutile | *de (6)* |

Une solution pour le paradoxe de l'analyse qui résulte notamment des idées émises par Gottlob Frege dans son essai *On Sense and Reference*, est la suivante. Cette solution remet en cause le passage de l'étape (2) à l'étape (3), qui conduit à la conclusion que l'analyse est triviale si l'*analysans* décrit exactement le contenu du concept α. Frege distingue en effet deux types de contenus sémantiques : d'une part, le sens ; et d'autre part, la référence. Dans ce contexte, il apparaît que si le concept α et son analysans ont la même *référence*, alors l'analyse qui en résulte est exacte. Toutefois, ceci n'interdit pas à l'analysans d'avoir un *sens* différent du concept α. Et dans de telles conditions, l'analyse se révèle pas triviale mais bien utile, par l'information nouvelle qu'elle procure.

# 40. Le problème de la rivière d'Héraclite



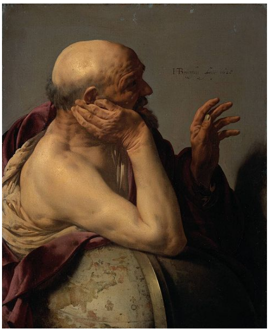
Le *problème de la rivière d'Héraclite* provient des *Fragments* de l'œuvre d'Héraclite qui sont parvenus jusqu'à nous. Héraclite y affirme qu'il n'est pas possible de traverser deux fois la même rivière, car les eaux qui constituent cette dernière sont constamment renouvelées. L'idée sous-jacente dans ce dernier problème est qu'entre deux traversées, la rivière a subi des changements tels qu'il ne s'agit plus exactement de la même rivière.

On peut formuler de manière plus précise le problème de la rivière d'Héraclite :

| | | |
|---|---|---|
| (1) | je traverse la rivière *r* au temps $T_1$ | *prémisse* |
| (2) | je traverse la rivière *r* au temps $T_2$ (avec $T_1 < T_2$) | *prémisse* |
| (3) | la rivière *r* a subi des changements entre $T_1$ et $T_2$ | *prémisse* |
| (4) | ∴ la rivière *r* au temps $T_1$ est différente de la rivière au temps $T_2$ | *de (3)* |
| (5) | ∴ au temps $T_2$ je traverse une rivière qui est différente de la rivière *r* que | *de (1),(2), (4)* |

| | | |
|---|---|---|
| | j'ai traversée au temps $T_1$ | |

Une objection qui a été formulée par rapport au problème de la rivière d'Héraclite est que les changements subis par la rivière entre $T_1$ et $T_2$ ne sont pas assez substantiels pour transformer la rivière en $T_1$ en une rivière différente en $T_2$. Selon ce point de vue, les changements subis par la rivière sont secondaires et n'affectent pas son identité en tant que rivière. Ce type d'objection, on le voit, a pour effet de bloquer le passage de l'étape (3) à l'étape (4). Il

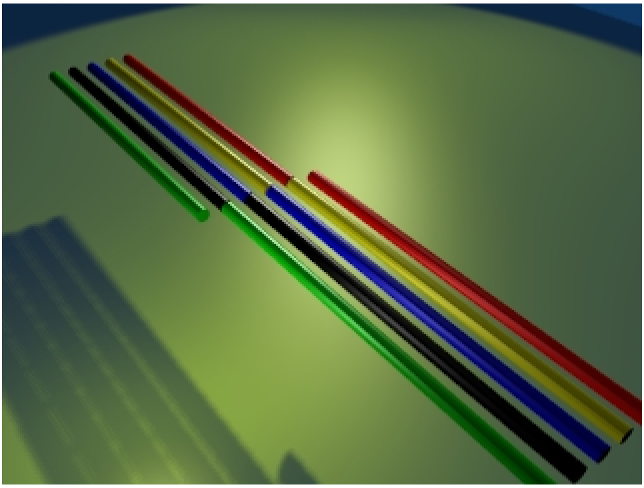


met ainsi l'accent sur la persistance de l'identité d'un objet *o* à travers le temps, malgré les changements de nature secondaire qui sont subis par cet objet. Car selon ce point de vue, les étapes (3) et (4) doivent être remplacées par :

| | | |
|---|---|---|
| (3*) | la rivière *r* a subi des changements mineurs entre $T_1$ et $T_2$ | *prémisse* |
| (4*) | $\therefore$ la rivière *r* au temps $T_1$ n'est pas différente de la rivière *r* au temps $T_2$ | *de (3*)* |

Pourtant, une telle objection ne suffit à résoudre définitivement le problème de la rivière d'Héraclite. En effet, la distinction sous-jacente entre les changements substantiels ou non-substantiels qui peuvent affecter un objet donné, se révèle difficile à appliquer. Ainsi, entre deux positions temporelles données, l'eau de la rivière a été entièrement renouvelée, de sorte que les éléments qui composent cette dernière ont été entièrement changés. Il est difficile alors de considérer que la totalité des éléments qui composent un objet à un moment donné ne constituent pas des éléments essentiels de celui-ci.

## Conclusion



Les paradoxes, arguments et problèmes philosophiques qui ont été exposés dans les pages précédentes ne constituent qu'une sélection parmi les nombreux problèmes abordés dans la riche littérature qui constitue la philosophie analytique contemporaine. Car il s'agit là d'un domaine vivant et évolutif, où chaque année, de nouveaux arguments voient le jour, sont ensuite exposés, puis discutés. On a pu le constater, des paradoxes millénaires non résolus y côtoient des arguments philosophiques qui viennent tout juste d'être décrits.

D'autre part, la présentation de ces problèmes contemporains de philosophie analytique a surtout pour but de permettre une meilleure connaissance du style analytique au lecteur qui est davantage familier avec la philosophie dite continentale. Car les deux styles, on l'a vu, constituent deux facettes de la philosophie, qui méritent toutes deux la respectabilité. L'objectif a simplement été ici de présenter une facette souvent méconnue de la philosophie contemporaine. Certains se sentiront d'emblée une affinité naturelle avec le style analytique. D'autres lui préféreront le style « continental » auquel ils sont attachés. Tous cependant, je l'espère, tireront profit d'une meilleure connaissance de la diversité des styles philosophiques.

De l'exposé des paradoxes et arguments qui précèdent, il ressort également, je le crois, que le raisonnement humain s'avère perfectible et étonnamment vulnérable à l'erreur. Car les pièges du raisonnement qui ont été décrits, les contradictions auxquelles nous entraînent aisément les paradoxes, indiquent que notre façon de raisonner à tous se révèle vulnérable. Il est assez fascinant de constater à quel point nous sommes tous enclins à raisonner d'une manière qui conduit à des conclusions paradoxales, nous laissant avec les contradictions qui résultent d'un raisonnement qui paraissait pourtant tout à fait valide. Le raisonnement qui conduit à l'erreur nous est commun, et là encore, si une solution devait être apportée à tel ou tel problème ou paradoxe, elle devrait pour être validée, se révéler consensuelle. On le voit, un tel domaine possède une portée pratique considérable. Il s'agit là d'améliorer et de perfectionner le mode de raisonnement qui nous est commun. Dans ce contexte, la découverte d'une solution consensuelle pour tel ou tel argument ou paradoxe non résolu devrait ainsi bénéficier à tous.

# Bibliographie


Arntzenius, F. & McCarthy, D. (1997) The two envelope paradox and infinite expectations, *Analysis*, 57, 42-50 [Le paradoxe des deux enveloppes]

Arntzenius, F. (2002) Reflections on Sleeping Beauty, *Analysis*, 62-1, 53-62 [Le problème de la Belle au Bois Dormant]

Axelrod, R. (1984) *The Evolution of Cooperation*, New York: Basic Books [Le dilemme du prisonnier]

Ayer A. J. (1976) *Language, Truth, and Logic*, Harmondsworth: Penguin, traduction Okana J. (1956) *Langage, vérité et logique*, Paris: Flammarion.

Ayer, A. J. (1973) On a Supposed Antinomy, *Mind*, 82, 125-126 [Le paradoxe de l'examen-surprise]

Barrow, J.D., and Tipler, F.J. (1986) *The Anthropic Cosmological Principle*, Oxford: Clarendon Press

Barwise, J. & Etchemendy, J. (1987) *The Liar: An Essay in Truth and Circularity*, Oxford University Press [Le paradoxe du Menteur]

Beaney, Michael, Analysis, *The Stanford Encyclopedia of Philosophy* (Summer 2003 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/sum2003/entries/analysis [Le paradoxe de l'analyse]

Binkley, R. (1968) The Surprise Examination in Modal Logic, *Journal of Philosophy*, 65, 127-136 [Le paradoxe de l'examen-surprise]

Bostrom, N. (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, Routledge, New York [L'argument de l'Apocalypse]

Bostrom, N. (2003) Are You Living In a Computer Simulation?, *Philosophical Quarterly*, 53, 243-255 [L'argument de la simulation]

Bovens, L. (1997) P and I Will Believe that not-P: Diachronic Constraints on Rational Belief, *Mind*, 104, 737-760

Broome, J. (1995) The two-envelope paradox, *Analysis*, 55, 6-11 [Le paradoxe des deux enveloppes]

Brueckner, A. (1986) Brains in a Vat, *Journal of Philosophy*, 83, 148-67

Burge, T. (1979) *Individualism and the Mental*, dans French, et al., eds. Midwest Studies in Philosophy: Studies in Metaphysics, Minneapolis: University of Minnesota Press

Burge, T. (1988) Individualism and self-knowledge, *Journal of Philosophy*, 85, 649-663

Campbell, R., and L. Sowden, éd. (1985) *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*, Vancouver: University of British Columbia Press [Le dilemme du prisonnier] [Le problème de Newcomb]

Carroll L. (1895) What The Tortoise Said To Achilles, *Mind*, 4, 278-280 [L'argument de Lewis Carroll]

Chow, T. Y. (1998) The Surprise Examination or Unexpected Hanging Paradox, *American Mathematical Monthly*, 105, 41-51 [Le paradoxe de l'examen-surprise]

Clark, M. & Shackel, N. (2000) The two-envelope paradox, *Mind*, 109, 415-42 [Le paradoxe des deux enveloppes]

Delahaye, J.-P., (2003) La Belle au bois dormant, la fin du monde et les extraterrestres, *Pour la Science*, 309, 98-103 [Le problème de la Belle au Bois Dormant]

Delahaye, J-P. (2004) Les inattendus mathématiques, Paris: Pour la Science/Belin [Les paradoxes et problèmes liés aux mathématiques]

Descartes, R. (2000) *Discours de la méthode*, Paris: Flammarion

Descartes, R. (1997) *Méditations métaphysiques*, Paris: Flammarion

Deutsch, H., (2002) Relative Identity, *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), E. N. Zalta (ed.),

http://plato.stanford.edu/archives/sum2002/entries/identity-relative/
Dietl, P. (1973) The Surprise Examination, *Educational Theory*, 23, 153-158 [Le paradoxe de l'examen-surprise]
Drange, T. M. (2000) The Fine-Tuning Argument Revisited, *Philo* 3, 2:45
Eckhardt, W. (1993) Probability Theory and the Doomsday Argument, *Mind*, 102, 483-488 [L'argument de l'Apocalypse]
Elga, A. (2000) Self-locating Belief and the Sleeping Beauty Problem, *Analysis*, 60-2, 143-147 [Le problème de la Belle au Bois Dormant]
Engel, P. (1994) Introduction à la philosophie de l'esprit, Paris: La Découverte [Les paradoxes et problèmes liés à la philosophie de l'esprit]
Engel, P. (1997) *La Dispute*, Minuit [Une introduction à la philosophie analytique]
Ewing A. C. (1962) *The fundamental questions of philosophy*, Collier Books: New York
Feyerabend, P. (1968) A Note on two 'Problems' of Induction , *British Journal for the Philosophy of Science*, 19, 251-253 [Le paradoxe de Goodman]
Franceschi, P. (1999) Comment l'urne de Carter et Leslie se déverse dans celle de Hempel, *Canadian Journal of Philosophy*, 29, 139-156 [L'argument de l'Apocalypse] [Le paradoxe de Hempel]
Franceschi, P. (2001) Une solution pour le paradoxe de Goodman, *Dialogue*, 40, 99-123 [Le paradoxe de Goodman]
Frege, G. (1892) *On Sense and Reference*, P. Geach et M. Black (éds.), Translations from the Philosophical Writings of Gottlob Frege. Oxford: Blackwell, 1960 [Le paradoxe de l'analyse]
Gettier E. L. (1963) Is Justified True Belief Knowledge?, *Analysis*, 23, 121-123 [Le problème de Gettier]
Goguen, J.A. (1969) The logic of inexact concepts,

*Synthese*, 19, 325-373 [Le paradoxe sorite]
Goldstein, L. (2003) Farewell to Grelling, *Analysis*, 63, 31-32 [Le paradoxe de Grelling]
Goodman, N. (1946) A Query On Confirmation, *Journal of Philosophy*, 43, 383-385, dans *Problems and Projects*, Indianapolis, Bobbs-Merrill, 1972, 363-366 [Le paradoxe de Goodman]
Goodman, N. (1984) *Fact, Fiction and Forecast* (1954), Cambridge, MA: Harvard University Press, traduction Abran M. (1984) *Faits, fictions et prédictions*, Paris: Editions de Minuit [Le paradoxe de Goodman]
Gott, R. (1993) Implications of the Copernican Principle for our Future Prospects, *Nature,* 363, 315-319 [L'argument de l'Apocalypse]
Hacking, I. (1993) *Le plus pur nominalisme*, Combas, L'éclat [Le paradoxe de Goodman]
Hájek, A. (2001) Pascal's Wager, *The Stanford Encyclopedia of Philosophy* (Winter 2001 Edition), E. N. Zalta (éd.), http://plato.stanford.edu/archives/win2001/entries/pascal-wager/ [Le pari de Pascal]
Hauser, L. (2001) The Chinese Room Argument, *The Internet Encyclopedia of Philosophy*, http://www.utm.edu/research/iep/ [L'argument de la Chambre chinoise]
Hempel, C. (1945) Studies in the logic of confirmation, *Mind*, 54, 1-26 et 97-121 [Le paradoxe de Hempel]
Huggett, N. (2002) Zeno's Paradoxes, *The Stanford Encyclopedia of Philosophy* (Fall 2002 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/fall2002/entries/paradox-zeno/ [Le paradoxe de la course]
Hughes, C. (1997) Same-kind coincidence and the ship of Theseus, *Mind*, 106, 53-69 [Le problème de Thésée]
Hume, D. (1779) *Dialogues concerning Natural Religion,* London, traduction Malherbe M. (1973) *Dialogues sur la religion naturelle*, Paris: Vrin
Hyde, D. (2002), Sorites Paradox, *The Stanford*

*Encyclopedia of Philosophy*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2002/entries/sorites-paradox [Le paradoxe sorite]

Irvine, A. D. (2003) Russell's Paradox, The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/sum2003/entries/russell-paradox/ [Le paradoxe de Russell]

Jackson, F. (1975) Grue, *Journal of Philosophy*, 72, 113-131 [Le paradoxe de Goodman]

Jackson, F., Menzies, P. & Oppy, G. (1994) The Two Envelope « Paradox », *Analysis*, 54, 43-45 [Le paradoxe des deux enveloppes]

Janaway, C. (1989) Knowing About Surprises: A Supposed Antinomy Revisited, *Mind*, 98, 391-410 [Le paradoxe de l'examen-surprise]

Jeffrey, R. C. (1983) *The Logic of Decision*, 2nd edition, University of Chicago Press

Kant E. (1787), *Critique de la raison pure*, Paris: Aubier, 1997

Kripke, S. (1975) Outline of a Theory of Truth, *Journal of Philosophy*, 72, 690-716 [Le paradoxe du Menteur]

Lassègue, J. (1996) What Kind of Turing Test did Turing have in Mind?, *Tekhnema : Journal of Philosophy and Technology*, 3, 37-58 [Le test de Turing]

Leslie, J. (1992) Time and the Anthropic Principle, *Mind*, 101, 521-540 [L'argument de l'Apocalypse]

Leslie, J. (1993) Doom and Probabilities, *Mind*, 102, 489-491 [L'argument de l'Apocalypse]

Leslie, J. (1996) *The End of the World: the science and ethics of human extinction*, London, Routledge [L'argument de l'Apocalypse]

Lewis, D. (1979) Prisoner's Dilemma Is a Newcomb Problem, *Philosophy and Public Affairs*, 8, 235-240 [Le dilemme du prisonnier] [Le problème de Newcomb]

Lewis, D. (2001) Sleeping Beauty: Reply to Elga,

*Analysis*, 61-3, 171-176 [Le problème de la Belle au Bois Dormant]
Malcolm, N. (1960) Anselm's Ontological Arguments, *Philosophical Review*, 69, 41-62
Martin, R. (1984) *Recent Essays on Truth and the Liar Paradox*, Oxford University Press [Le paradoxe du Menteur]
McClennen, E. (1994) *Finite Decision Theory*, dans Jordan, J., éd.. (1994) *Gambling on God: Essays on Pascal's Wager*, Rowman & Littlefield
McGee, V. (1991), *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*, Hackett Publishing [Le paradoxe du Menteur]
McGrew T. & McGrew L. & Vestrup E. (2001) Probabilities and the Fine-Tuning Argument: a Sceptical View, *Mind*, 110, 1027-1038
McGrew, T., Shier, D. & Silverstein, H. (1997) The two envelope paradox resolved, *Analysis*, 57, 28-33 [Le paradoxe des deux enveloppes]
McTaggart, J. E. (1908) The Unreality of Time, *Mind*, 457-474 [L'argument de McTaggart]
Monton, B. (2002) Sleeping Beauty and the Forgetful Bayesian, *Analysis*, 62-1, 47-53 [Le problème de la Belle au Bois Dormant]
Moore, G.E. (1942) A reply to my critics, in *The Philosophy of G.E. Moore* , éd. P.Schlipp 543667, Evanston: Tudor
Nozick, R. (1969) *Newcomb's problem and two principles of choice*, dans N. Rescher, éd., Essays in Honor of Carl G. Hempel, Dordrecht: Reidel, 114-146 [Le problème de Newcomb]
O' Connor, D. (1948) Pragmatic paradoxes, *Mind*, 57, 358-359 [Le paradoxe de l'examen-surprise]
Oppy, G. & Dowe, D. (2003) The Turing Test, The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/sum2003/entries/turi

ng-test/ [Le test de Turing]
Oppy, G. (1990) On Rescher on Pascal's Wager, *International Journal for Philosophy of Religion*, 30, 159-68 [Le pari de Pascal]
Oppy, G. (1995) *Ontological Arguments and Belief in God*, New York: Cambridge University Press [L'argument ontologique]
Oppy, G. (2002) Ontological Arguments, *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), E. N. Zalta (éd.), http://plato.stanford.edu/archives/sum2002/entries/ontological-arguments/ [L'argument ontologique]
Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press
Pascal, B. (2004) *Les Pensées*, Paris: Gallimard [Le pari de Pascal]
Pettit, P. & Sugden, R. (1989) The Backward Induction Paradox, *Journal of Philosophy*, 86, 169-182
Plantinga, A. (1967) *God and Other Minds - A Study of Rational Justification of Belief in God*, Ithica, New York: Cornell University Press
Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press
Platon, *La République*, http://www.cvm.qc.ca/encephi/menus/textes.htm
Poundstone, W. (1990) *Les labyrinthes de la raison*, Belfond [Les paradoxes en général]
Putnam, H. (1975) The Meaning of 'Meaning', dans *Mind, Language, and Reality*, Cambridge: Cambridge University Press, 215-271
Putnam, H. (1982) *Reason, Truth and History*, Cambridge: Cambridge University Press, traduction Gerschenfeld A. (1984) *Raison, vérité et histoire,* Paris: Minuit
Quine, W. (1953) On a So-called Paradox, *Mind*, 62, 65-66 [Le paradoxe de l'examen-surprise]
Rosenthal, D.M. (2002) Moore's paradox and Crimmins's

case, *Analysis*, 62, 167171 [Le paradoxe de Moore]
Ross, H. (1998) *Big Bang Refined by Fire*, Pasadena, CA: Reasons to Believe
Russell, B. (1923) Vagueness, *The Australian Journal of Philosophy and Psychology*, 1, 84-92 [Le paradoxe sorite]
Russell, B. (1980) *Correspondence with Frege*, dans Philosophical and Mathematical Correspondence, by Gottlob Frege, traduction par Hans Kaal, Chicago: University of Chicago Press
Sainsbury, M. (1995) (2ème éd.) *Paradoxes*, Cambridge: Cambridge University Press [Les paradoxes en général]
Saint Anselme (1077), *Proslogion*, trad. B. Pautrat, éd. Garnier Flammarion, 1993 [L'argument ontologique]
Salmon, N. (1986) *Frege's Puzzle*, Cambridge MA: MIT Press
Salmon, W. C. (ed.) (1970) *Zeno's Paradoxes*, Indianapolis & New York: Bobbs-Merrill [Le paradoxe de la course]
Savage, W. (1967) The paradox of the stone, *Philosophical Review*, 76, 74-79 [Le paradoxe de la pierre]
Saygin, A.P., Cicekli, I. & Akman, V. (2000), Turing Test: 50 Years Later, *Minds and Machines*, 10(4), 463-518 [Le test de Turing]
Schrader, D. (1979) A solution to the stone paradox, *Synthese*, 42, 255-264 [Le paradoxe de la pierre]
Scriven, M. (1951) Paradoxical announcements, *Mind*, 60, 403-407 [Le paradoxe de l'examen-surprise]
Searle, J. (1980) Minds, brains, and programs, *Behavioral and Brain Sciences*, 3, 417-424
Searle, J. (1984) *Minds, Brains, and Science*, Cambridge: Harvard University Press
Searle, J. (1990) Is the Brain's Mind a Computer Program?, *Scientific American*, 262, 26-31
Smith, J. W. (1984) The surprise examination on the

paradox of the heap, *Philosophical Papers*, 13, 43-56 [Le paradoxe de l'examen-surprise] [Le paradoxe sorite]
Smith, Q. & Oaklander, L. N. (1994) *The New Theory of Time*, Yale University Press, New Haven
Sorensen, R. A. (1982) Recalcitrant versions of the prediction paradox, *Australasian Journal of Philosophy*, 69, 355-362 [Le paradoxe de l'examen-surprise]
Sorensen, R. A. (1988) *Blindspots*, Oxford: Clarendon Press [Le paradoxe sorite] [Le paradoxe de l'examen-surprise]
Sorensen, R.A. (2000) Moore's problem with iterated belief, *Philosophical Quarterly*, 50, 2843 [Le paradoxe de Moore]
Sowers, G. F. (2002) The Demise of the Doomsday Argument, *Mind*, 111, 37-45 [L'argument de l'Apocalypse]
Stroud, B. (1989) *The Significance of Philosophical Scepticism*, Oxford: Oxford University Press
Sullivan, P. (2003) A note on incompleteness and heterologicality, *Analysis*, 63, 32-38
Suppes, P. (1972) *Axiomatic Set Theory*, New York: Dover
Swinburne, R. (1968) The Argument from Design, *Philosophy*, 43, 199-211
Turing, A. M. (1950) Computing Machinery and Intelligence, *Mind*, 59, 433-460 [Le test de Turing]
Ullian, J. S. (1961) More one 'Grue' and Grue, *Philosophical Review*, 70, 386-389 [Le paradoxe de Goodman]
Vidal-Rosset, J. (2004) *Qu'est-ce qu'un paradoxe?*, Paris: Vrin [Les paradoxes en général]
Weatherson, B. (2004) Are You a Sim?, *Philosophical Quarterly*, 53, 425-431 [L'argument de la simulation]
Wiggins, D. (1982) *Heraclitus' Conceptions of Flux, Fire, and Material Persistence*, dans Schofield, M. &

Nussbaum, M., éds., Language and Logos, Cambridge University Press

Williams, J.N. (1979) Moore's paradox: One or Two, *Analysis*, 39, 141-142 [Le paradoxe de Moore]

Williams, J.N. (1999) Wittgensteinian accounts of Moorean absurdity, *Philosophical Studies,* 92, 283306

Williamson, T. (1994) *Vagueness*. London: Routledge [Le paradoxe sorite]

Zadeh, L. (1975) Fuzzy logic and approximate reasoning, *Synthese*, 30, 407-428 [Le paradoxe sorite]

Zalta, E. N. (2002) Gottlob Frege, *The Stanford Encyclopedia of Philosophy* (Spring 2002 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/spr2002/entries/frege/ [Le problème de Frege relatif aux propositions d'identité]

## Sites Internet


Adam Elga, http://www.princeton.edu/~adame/, site d'auteur

Brian Weatherson, http://brian.weatherson.net/papers.html, site d'auteur

Descartes, *Discours de la méthode*, http://abu.cnam.fr/BIB/auteurs/descartesr.html

Descartes, *Les méditations métaphysiques*, http://abu.cnam.fr/BIB/auteurs/descartesr.html

Eliott Sober, http://philosophy.wisc.edu/sober/papers.htm, site d'auteur

Graham Oppy, http://www.arts.monash.edu.au/phil/department/Oppy/grahampapers.html, site d'auteur

Institut Jean Nicod, http://jeannicod.ccsd.cnrs.fr/, site contenant les articles des chercheurs de l'Institut Jean Nicod

Nicholas J. J. Smith, http://www.personal.usyd.edu.au/~njjsmith/papers/, site d'auteur

Pascal, *Les Pensées*, http://abu.cnam.fr/BIB/auteurs/pascalb.html

Roy Sorensen http://www.dartmouth.edu/~rasoren/papers/papers.html, site d'auteur

The Anthropic Principle, de Nick Bostrom, http://www.anthropic-principle.com, site Internet relatif à l'argument de l'Apocalypse et au problème de la Belle au bois dormant

The Internet Encyclopedia of Philosophy, http://www.utm.edu/research/iep/, encyclopédie philosophique

The Simulation Argument, de Nick Bostrom, http://www.simulation-argument.com, site Internet relatif à l'argument de la simulation

The Stanford Encyclopedia of Philosophy, éditée par E. N. Zalta, http://plato.stanford.edu/archives/, encyclopédie philosophique

Paul Franceschi, http://www.univ-corse.fr/~franceschi, le site de l'auteur

# REMERCIEMENTS

Je remercie Francis Antona et Christian Carayon pour des commentaires très utiles pendant la rédaction du présent ouvrage.



# CRÉDITS

Les illustrations ont été réalisées à l'aide du logiciel Blender (http://www.blender.org/).

# TABLE DES MATIÈRES

## Table des matières

**INTRODUCTION....................................................................5**

**1. LE PARADOXE DU MENTEUR.....................................11**

**2. LE PARADOXE SORITE.................................................15**

**3. LE PARADOXE DE RUSSELL.......................................21**

**4. LE PARADOXE DE L'EXAMEN-SURPRISE...............25**

**5. LE PARADOXE DE GOODMAN....................................31**

**6. LE PROBLÈME DE NEWCOMB....................................37**

**7. LE DILEMME DU PRISONNIER ...................................41**

**8. LE PARADOXE DE CANTOR........................................45**

**9. LE PARADOXE DE GRELLING....................................49**

**10. LE PARADOXE DES DEUX ENVELOPPES ..............53**

**11. LE PARADOXE DE MOORE........................................57**

**12. LE PARADOXE DE LÖB...............................................61**

**13. LE PARADOXE DE LA COURSE................................65**

**14. LE PARADOXE DE LA PIERRE..................................69**

**15. L'ARGUMENT DE L'APOCALYPSE..........................73**

**16. LE PROBLÈME DU NAVIRE DE THÉSÉE................79**

**17. LE PROBLÈME DE HEMPEL......................................83**

**18. L'ARGUMENT DE MCTAGGART..............................87**

**19. L'ARGUMENT ONTOLOGIQUE................................91**

**20. L'ARGUMENT DU RÉGLAGE OPTIMAL.................95**

**21. L'ARGUMENT DU RÊVE............................................99**

**22. L'EXPÉRIENCE DES « CERVEAUX DANS UNE CUVE » ..........................................................................103**

**23. L'ARGUMENT TÉLÉOLOGIQUE.............................107**

**24. L'ARGUMENT DU PARI DE PASCAL.......................111**

**25. L'ARGUMENT SELON LE MAL...............................117**

**26. LE COGITO CARTÉSIEN..........................................121**

**27. L'ARGUMENT DE LEWIS CAROLL........................125**

**28. L'EXPÉRIENCE DE PENSÉE DE LA TERRE JUMELLE........................................................................129**

**29. L'ARGUMENT CONTRE LE PRINCIPE DE VÉRIFIABILITÉ.................................................................133**

**30. L'ALLÉGORIE DE LA CAVERNE...........................137**

**31. L'ARGUMENT DE LA SIMULATION......................145**

**32. L'ARGUMENT DUALISTE EN VERTU DE LA DIVISIBILITÉ...................................................................149**

**33. LE PROBLÈME DE LA BELLE AU BOIS DORMANT ..........................................................................................153**

**34. L'ARGUMENT DU MAUVAIS GÉNIE......................157**

35. L'ARGUMENT DE LA CHAMBRE CHINOISE DE SEARLE....................................................................161

36. LE TEST DE TURING ................................................165

37. LE PROBLÈME DE GETTIER..................................169

38. LE PROBLÈME DE FREGE RELATIF AUX PROPOSITIONS D'IDENTITÉ........................................173

39. LE PARADOXE DE L'ANALYSE...............................177

40. LE PROBLÈME DE LA RIVIÈRE D'HÉRACLITE. 179

CONCLUSION....................................................................181

BIBLIOGRAPHIE..............................................................183

SITES INTERNET.............................................................193

CRÉDITS.............................................................................195

TABLE DES MATIÈRES..................................................196