

Introduction

Paul Bernier

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004901ar>
DOI : <https://doi.org/10.7202/004901ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce document

Bernier, P. (2000). Introduction. *Philosophiques*, 27 (1), 3–10.
<https://doi.org/10.7202/004901ar>

Introduction

PAUL BERNIER

bernierp@umoncton.ca

En philosophie contemporaine, la thèse du monisme matérialiste est très largement acceptée comme étant un *desideratum* général que devrait satisfaire une théorie de l'esprit. En son sens le plus large, cette thèse dit simplement que, d'un point de vue ontologique, il n'existe d'autre substance que la substance matérielle. Mais dès lors que l'on souscrit à cette thèse, la tentation est presque irrésistible, du moins à première vue, de faire un pas additionnel et d'endosser l'idée d'une réduction psychophysique, c'est-à-dire d'une réduction des états mentaux à des états physiques, ceux du cerveau, comme le suggère notamment la thèse de l'identité type-type¹ ou encore le modèle classique de la réduction nagelienne², appliquée au problème corps/esprit. Mais, comme certains des auteurs qui contribuent à ce numéro le font remarquer, c'est la doctrine du physicalisme non réductionniste qui a constitué l'orthodoxie en philosophie de l'esprit au cours des trente dernières années³. Selon cette doctrine, même si tous les objets que nous devons admettre dans notre ontologie sont des objets constitués uniquement de matière, les propriétés mentales seraient irréductibles à des propriétés physiques, et elles

1. Pour les versions standard de la thèse de l'identité type-type, voir Smart, J. J. C., « Sensations and Brain Processes », *Philosophical Review*, 68, 1959, p. 141-156 ; Place, U. T., « Is Consciousness a Brain Process? », dans Lycan, W. G., dir., *Mind and Cognition*, 2^e édition, Oxford, Blackwell, 1999, p. 14-19 (cet article fut publié originalement en 1956) ; Feigl, H., « The "Mental" and the "Physical" », *Minnesota Studies in the Philosophy of Science*, Minneapolis, University of Minnesota Press, 1958, p. 370-497.

2. La réduction nagelienne est un modèle général visant à élucider la structure logique de la réduction interthéorique ; voir Nagel, E., *The Structure of Science*, New York, Harcourt, Brace and World, 1961 ; voir également la contribution de Kim au présent numéro, ainsi que celles de Kistler et de Poirier. Par ailleurs, il est notoire que les tenants du behaviorisme logique souscrivaient également au monisme matérialiste. Même si le behaviorisme logique a été généralement rejeté, l'approche fonctionnaliste qui a été prédominante en théorie de l'esprit au cours des trente dernières années s'est inspirée des analyses dispositionnalistes de cette doctrine. Pour les versions les plus influentes du fonctionnalisme, voir Putnam, H., « The Nature of Mental States », dans Putnam, H., *Mind, Language, and Reality*, Cambridge (Mass.), Cambridge University Press, 1975 (originalement paru en 1967 sous le titre « Psychological Predicates ») ; Armstrong, D., *A Materialist Theory of Mind*, London, Routledge & Kegan Paul, 1968 ; et Lewis, D., « Psychophysical and Theoretical Identifications », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

3. Pour les principaux arguments ayant contribué à l'établissement du physicalisme non réductionniste comme l'orthodoxie en théorie de l'esprit, voir Putnam, H., « The Nature of Mental States » et Davidson, D., « Mental Events », dans Foster, L. et Swanson, J. W., dir., *Experience and Theory*, Amherst (Mass.), University of Massachusetts Press, 1970, et Fodor, J. A., « Special Sciences », dans Fodor, J. A., *Representations*, Cambridge (Mass.), Cambridge University Press, 1981. Pour une discussion du physicalisme non réductionniste, voir particulièrement ici les contributions de Kim, Garrett, et Poirier.

formeraient un domaine d'investigation, celui de la psychologie, qui serait autonome relativement aux autres domaines scientifiques. Dans la littérature récente, le physicalisme non réductionniste a été très largement associé à la thèse fonctionnaliste, et plus précisément à son corrélat, soit la thèse de la réalisation multiple, qui pour plusieurs constituerait une réfutation du principe même de la possibilité du réductionnisme. Selon ce point de vue, une propriété mentale étant identifiée à un certain rôle fonctionnel, et un rôle fonctionnel pouvant être réalisé, physiquement, de multiples façons qui sont physiquement hétérogènes, une propriété mentale ne pourrait être réduite à une propriété physique⁴.

Jaegwon Kim est certainement l'un des philosophes contemporains qui a le plus contribué à clarifier notre compréhension du physicalisme et de ses enjeux⁵. Et il est sans doute celui qui a le plus vigoureusement critiqué la doctrine du physicalisme non réductionniste, et le fonctionnalisme, du moins lorsque ce dernier est interprété dans sa version non réductionniste, sur la base de la thèse de la réalisabilité multiple. Dans son article, Kim met en lumière une intuition importante liée au réductionnisme psychophysique, qui a eu tendance à être complètement négligée dans la littérature récente. Cette intuition était au cœur du problème que les émergentistes britanniques du début du siècle avaient soulevé relativement au réductionnisme psychophysique. Selon cette intuition émergentiste, un modèle adéquat de la réduction devrait satisfaire un certain *desideratum*, et Kim défend la thèse selon laquelle, si la réduction psychophysique est entendue dans le sens de la réduction nagelienne classique — comme elle l'a généralement été dans les discussions récentes en philosophie de l'esprit —, elle serait incapable de satisfaire ce *desideratum*. Sur cette base, Kim propose ensuite un autre modèle de la réduction, qu'il appelle la « réduction fonctionnelle » et qui, selon lui, ne serait pas confronté à cette difficulté, dans la mesure où il arriverait à satisfaire le *desideratum* réductionniste. Kim suggère enfin qu'il est tout simplement incohérent de tenir la thèse fonctionnaliste comme étant une version du physicalisme non réductionniste, ce qui est manifeste dès lors que nous devons préférer le modèle de la réduction fonctionnelle à celui de la réduc-

4. Ce ne sont pas toutes les versions importantes du fonctionnalisme qui conçoivent cette théorie comme étant incompatible avec le réductionnisme. Les versions défendues par David Armstrong et David Lewis sont des exceptions notoires. À ce sujet, voir l'article de Kim qui souligne qu'il y aurait, selon lui, une incohérence dans l'idée que la thèse de la réalisation multiple entraînerait une forme d'antiréductionnisme. Voir également l'article de Poirier.

5. Nous ne pouvons proposer ici une bibliographie des très nombreux articles et ouvrages que Kim a rédigés sur cette question, mais il est utile de noter quelques titres récents : Kim, J., « Multiple Realization and the Metaphysics of Reduction », *Philosophy and Phenomenological Research*, 52, 1992 ; Kim, J., *Supervenience and Mind. Selected Philosophical Essays*, Cambridge (Mass.), Cambridge University Press, 1993 ; Kim, J., « The Mind-Body Problem : Taking Stock After Forty Years », dans Tomberlin, J., dir., *Philosophical Perspectives*, 11, Boston, Blackwell, 1997 ; Kim, J., *Mind in a Physical World*, Cambridge (Mass.), MIT Press, 1998.

tion nagelienne⁶. Cependant, bien qu'il soit d'avis que la plupart des états mentaux peuvent être réduits, au sens de la réduction fonctionnelle, Kim conclut en indiquant qu'il croit que les *qualia*, ou le caractère phénoménal de l'expérience sensorielle, ne peuvent être réduits.

Dans son commentaire critique du texte de Kim, Max Kistler soulève deux objections à l'argument de Kim. Premièrement, il soutient que si l'on devait accepter le *desideratum* émergentiste, sur la base duquel Kim rejette le modèle de la réduction nagelienne, alors on devrait rejeter certaines réductions standard qui sont admises comme des exemples de réduction qui ne posent pas de problème. Deuxièmement, il propose que si effectivement nous devons admettre le *desideratum* émergentiste, alors il y a des raisons de croire que le modèle de la réduction fonctionnelle ne serait pas dans une meilleure position que celui de la réduction nagelienne, dans la mesure où, comme ce dernier, la réduction fonctionnelle devrait également reposer sur des lois de correspondance, et que celles-ci doivent être empiriques et non pas une question de définition *a priori*.

Un aspect important du point de vue général de Kim consiste à réfuter l'argument antiréductionniste qui s'appuie sur la thèse de la réalisabilité multiple. Pierre Poirier reprend cet aspect du point de vue de Kim, en le développant et en l'appuyant sur des considérations concernant la méthodologie aussi bien en physique qu'en biologie. Poirier distingue deux versions que cet argument antiréductionniste peut prendre, soit une version *interspécifique* et une version *intraspécifique*, et il soutient que ni l'une ni l'autre ne permettent de tirer la conclusion antiréductionniste. La version interspécifique de l'argument est rejetée sur la base d'un principe de relativisation non *ad hoc* de la réduction à des domaines bien définis, comme en fait foi le concept macroscopique de température dont la réalisabilité multiple ne bloque aucunement la réductibilité. De plus, Poirier développe une suggestion récente de John Bickle⁷ selon laquelle le cas de la température fournirait également une analogie avec la réalisation multiple *intraspécifique*, dans la mesure où deux gaz peuvent être de même température, même si à un certain niveau de description assez fin ils sont réalisés différemment. Sur cette base, Poirier suggère une certaine façon de concevoir la réduction comme faisant appel à des idéalizations mathématiques se situant à un certain niveau d'abstraction approprié, et il conclut en montrant que les neurosciences mathématiques permettent de mettre en évidence certaines propriétés abstraites de réseaux de neurones, qui se situent justement à un niveau approprié de généralité, c'est-à-dire à un niveau assez général pour que ces propriétés abstraites soient interprétées comme des propriétés macroscopiques ou « psychologiques ».

6. Voir le texte de Pierre Poirier qui défend également l'idée que le fonctionnalisme est une thèse réductionniste.

7. Voir Bickle, J., *Psychoneural Reduction*, Cambridge (Mass.), MIT Press, 1998.

Un des aspects du modèle de la réduction fonctionnelle proposé par Kim est qu'à la première étape de la réduction, on devrait donner une définition fonctionnelle de la propriété visée par la réduction, par exemple une propriété mentale. C'est-à-dire qu'on devrait dans un premier temps spécifier le rôle causal de cette propriété. C'est d'ailleurs cet aspect de la réduction fonctionnelle qui amène Kim à conclure son article en notant qu'il doute qu'une réduction fonctionnelle des *qualia* soit possible, compte tenu des sérieuses objections, désormais classiques, à la possibilité de définir fonctionnellement les *qualia*⁸. Dans sa contribution, Brian McLaughlin défend une nouvelle version de la thèse de l'identité type-type qui, depuis peu, commence à avoir une certaine influence⁹. L'argument de McLaughlin est subtil et rigoureusement développé, et il ne saurait être question de le résumer ici. Cela dit, l'article de McLaughlin tente de répondre à la question suivante : qu'est-ce qu'un démon de Laplace devrait connaître s'il devait être en mesure de connaître tous les faits de notre monde, incluant ce que McLaughlin appelle des faits d'expérience, c'est-à-dire des faits relatifs aux *qualia*? Et il soutient qu'il est tout à fait possible de répondre à cette question d'une façon qui soit compatible avec le physicalisme. Le point central de son argumentation consiste à nier que, pour connaître tous les faits du monde actuel, incluant les faits d'expérience, un démon laplacéen devrait connaître des *lois phénoméno-physiques/fonctionnelles contingentes* établissant des corrélations entre les *qualia* et des états physiques correspondants. Si un démon laplacéen devait connaître de telles lois contingentes, cela laisserait ouverte la possibilité qu'il existe des mondes possibles physiquement identiques au monde actuel, qui seraient dépourvus de *qualia*. Mais McLaughlin soutient que pour connaître tous les faits du monde actuel, incluant les faits d'expérience, il suffirait au démon laplacéen de connaître, outre certains « faits physiques de base », certaines *identités* phénoméno-physiques/fonctionnelles, ce qui, comme il le propose, serait compatible avec le physicalisme, dans la mesure où de telles identités, bien qu'elles ne puissent être connues *qu'a posteriori*, n'en seraient pas moins nécessaires. Il est intéressant de noter que

8. Voir notamment Nagel, T., « What Is It Like To Be a Bat? », *Philosophical Review*, 83, 1974, p. 435-450 ; Block, N., « Troubles With Functionalism », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980 ; Jackson, F., « Epiphenomenal Qualia », *Philosophical Quarterly*, 32, 1982, p. 127-136 ; Shoemaker, S., « The Inverted Spectrum », *Journal of Philosophy*, 74, 1981, p. 357-381 ; Levine, J., « Materialism and Qualia : the Explanatory Gap », *Pacific Philosophical Quarterly*, 64, 1983, p. 354-361.

9. Voir Hill, C., « Imaginability, Conceivability, Possibility and the Mind-Body Problem », *Philosophical Studies*, 87, 1997, p. 61-85 ; Hill, C. et McLaughlin, B., « There are Fewer Things in Reality than Dreamt of in Chalmers' Philosophy », *Philosophy and Phenomenological Research*, 59, 1999, p. 445-454 ; Block, N. et Stalnaker, R., « Conceptual Analysis, Dualism, and the Explanatory Gap », à paraître dans *Philosophical Review* ; et McLaughlin, B., « In Defence of New Wave Materialism », dans Barry Loewer et Carl Gillett, dir., *Physicalism and Its Discontents*, Cambridge (Mass.), Cambridge University Press.

le point de vue développé par McLaughlin dans cet article est directement en conflit avec une thèse centrale du modèle de la réduction fonctionnelle de Kim, à savoir que pour être réductible une propriété mentale devrait être fonctionnalisable, c'est-à-dire qu'elle devrait avoir une définition fonctionnelle, ce qui correspond à la première étape du modèle de la réduction fonctionnelle.

Dans leurs contributions, Jérôme Dokic et Paul Bernier s'intéressent également au problème des *qualia*, ou de la conscience phénoménale. Plusieurs philosophes ont été tentés d'endosser une certaine stratégie influente, différente de celle de McLaughlin, pour essayer de résoudre le problème des *qualia*. Cette stratégie consiste à tenter de réduire les *qualia* à des états représentationnels, pour ensuite en rendre compte dans le cadre d'une théorie du contenu mental qui serait compatible avec le matérialisme¹⁰. Selon le point de vue développé par Jérôme Dokic, cette approche générale, qu'il appelle la Théorie Matérialiste Standard (ou TMS), serait confrontée à un sérieux problème dans la mesure où elle serait incapable de rendre compte de la différence, intuitivement plausible, entre ce que Dokic appelle « ressentir une douleur de l'intérieur » et « percevoir une douleur de l'extérieur ». Pour établir le bien fondé de cette distinction, et pour ensuite proposer une approche différente de la TMS qui soit tout de même compatible avec le matérialisme, Dokic se concentre sur le cas particulier de la conscience corporelle, et plus particulièrement sur l'expérience de la douleur. Une des conséquences de l'adoption de la TMS, qui conçoit les *qualia* non pas comme des propriétés intrinsèques de l'expérience, mais plutôt comme des propriétés intentionnelles présentées *dans* l'expérience — comme par exemple l'indication d'un dommage dans certains tissus du corps — est qu'il n'existerait pas de différence fondamentale entre la conscience phénoménale et l'expérience perceptive d'objets extérieurs. Dokic souligne très clairement les aspects importants que ces deux types d'expériences ont en commun, mais il soutient que, malgré ces similitudes, on doit les distinguer, et que c'est justement ce que la TMS serait incapable de faire. S'appuyant sur des arguments de David Pears et de Friedrich von Hayek, qu'il développe dans le contexte de sa discussion de la TMS des *qualia*, Dokic propose une certaine approche qui se veut mitoyenne entre, d'une part, une approche purement phénoménologique et, d'autre part, la TMS. L'approche suggérée est mitoyenne en ce sens que, contrairement à une approche purement phénoménologique, elle serait compatible avec le matérialisme et, contrairement à la TMS, elle arriverait à rendre compte de la différence entre « ressentir une douleur de l'intérieur » et la

10. Pour ce type de stratégie qui s'appuie sur la théorie de l'indication fiable, voir Dretske, E., *Naturalizing the Mind*, Cambridge (Mass.), MIT Press, 1995 ; et Tye, M., *Ten Problems of Consciousness*, Cambridge (Mass.), MIT Press, 1995 ; pour celle qui s'appuie sur une sémantique des rôles conceptuels, voir Harman, G., « The Intrinsic Quality of Experience », dans Tomberlin, J., dir., *Philosophical Perspectives*, vol. 4, Northridge (Calif.), Ridgeview, 1990.

« percevoir à l'extérieur ». Dokic conclut en proposant que ce qui distinguerait essentiellement la conscience des *qualia* corporels de la perception extérieure serait l'aspect réflexif de la première dont serait dépourvu la seconde.

La stratégie générale de naturalisation des *qualia* via leur réduction à des états intentionnels a également fait l'objet d'une critique importante de Ned Block, qui s'y est objecté en suggérant que, bien que nous devions accepter que les sensations aient un contenu intentionnel, elles auraient en outre, et de façon essentielle, un contenu purement phénoménal qui serait irréductible à leur contenu intentionnel. Dans notre propre contribution à ce numéro, nous discutons un argument de Block visant à établir l'existence de cette distinction qui, si elle était fondée, constituerait une objection dirimante à toute tentative de fonctionnalisation des *qualia*. Nous proposons une réfutation de l'argument de Block qui, si elle est correcte, appuierait une analyse fonctionnaliste du type de celle qui a été suggérée par Gilbert Harman. L'argument de Block, qui s'appuie sur l'expérience de pensée de la Terre inversée, s'inscrit dans la foulée des divers arguments antiphysicalistes qui ont été levés contre la possibilité d'une réduction fonctionnelle des *qualia*. Notre objectif, dans cet article est relativement restreint, dans la mesure où il se concentre exclusivement sur l'argument de Block ; mais s'il est correct, il montre que la distinction soumise par Block, entre le contenu phénoménal des sensations et leur contenu intentionnel, serait incohérente, à tout le moins dans le cadre du scénario de l'expérience de la Terre inversée qui est censé l'appuyer.

Pour sa part, Pierre Jacob discute deux problèmes concernant l'efficacité causale des états mentaux. Le premier est le problème du caractère extrinsèque du contenu intentionnel, qui découle directement de la thèse de l'externalisme du contenu, selon laquelle le contenu représentationnel dépendrait essentiellement de relations entre le sujet et son environnement. Le problème est alors de comprendre comment des états mentaux individués en accord avec l'externalisme, c'est-à-dire de façon non locale, pourraient avoir une efficacité causale dans la production du comportement, dans la mesure où l'origine causale d'un comportement semblerait dépendre uniquement de propriétés intrinsèques, ou non relationnelles des états mentaux. Dans la deuxième partie de son article, Jacob discute un problème plus général que Kim a abordé au sujet du physicalisme non réductionniste. Il s'agit du problème de l'exclusion explicative, qui repose sur un certain principe de parcimonie dans les explications causales. Ce problème se pose pour le physicalisme non réductionniste dans la version qui en est généralement admise, c'est-à-dire la version qui présuppose la thèse de l'identité token-token, selon laquelle une exemplification particulière d'un token mental serait identique à celle d'un token physique¹¹. Si on suppose qu'une exemplification d'une

11. Mais voir l'article de Brian Garrett pour une interprétation du physicalisme non réductionniste qui rejette la thèse de l'identité token-token.

certainne propriété mentale M est suivie (apparemment de façon causale) de celle d'une autre propriété mentale M^* , et que ces deux tokens mentaux sont respectivement nécessités par deux tokens physiques, P et P^* , et si l'on suppose qu'il existe une explication causale de la production de P^* par P , alors postuler l'existence d'un lien causal entre M et M^* semblerait superflu. Dans son article, Jacob propose des solutions à ces deux problèmes qui visent à défendre l'idée que le contenu intentionnel a bien une efficacité causale dans la production du comportement¹².

Dans ses critiques du physicalisme non réductionniste et du fonctionnalisme, Kim a mis en évidence un principe important, qui est justement au cœur du problème de l'exclusion explicative discutée par Jacob dans la deuxième partie de son texte. Il s'agit du principe de l'héritabilité causale. Selon ce principe, les pouvoirs causaux des propriétés mentales seraient hérités directement de ceux des événements physiques qui les exemplifient, dans la mesure où une propriété mentale est réalisée physiquement, et dans la mesure où les pouvoirs causaux d'une propriété mentale devraient relever uniquement de ceux des états physiques particuliers qui la réalisent, sous peine d'enfreindre le principe de la clôture physique du monde. Dans sa contribution, Brian Garrett développe un argument qui vise rien de moins que le rejet du principe d'héritabilité causale. Pour ce faire, il défend une certaine conception du physicalisme non réductionniste qui ferait l'économie de la thèse davidsonienne de l'identité token-token, thèse à laquelle le physicalisme non réductionniste est généralement associé. L'argument original de Garrett et le point de vue quelque peu audacieux qui en découle s'appuient sur des considérations métaphysiques selon lesquelles il serait tout à fait plausible qu'un particulier, bien qu'il soit constitué de certaines parties physiques, ne soit pas nécessairement identique à la somme méréologique des parties qui le composent. Comme le note Garrett, selon ce point de vue « la constitution n'est pas l'identité ». Cette idée, qui n'est pas nouvelle, a récemment été défendue et débattue dans le contexte général de la métaphysique des objets. L'originalité du point de vue de Garrett est d'appliquer à la notion d'événement cette idée métaphysique, qui connaît depuis peu un certain regain d'intérêt, et cela dans le but de tirer certaines conséquences sur le terrain de la métaphysique de l'esprit. En effet, lorsque appliquées aux événements, ces considérations permettent à Garrett de résister à la thèse de l'identité token-token et de produire un argument visant à réfuter le principe kimien de l'héritabilité causale. Le point de vue de Garrett paraîtra assez radical, dans la mesure où la thèse de l'identité token-token a pratiquement acquis le statut de dogme en théorie de l'esprit contemporaine. Mais Garrett

12. Les solutions proposées par Jacob s'inscrivent dans le cadre de l'approche qu'il a développée dans un livre récent. Voir Jacob, P., *What Minds Can Do*, Cambridge (Mass.), Cambridge University Press, 1997 ; et *id.*, *Pourquoi les choses ont-elles un sens?*, Paris, Odile Jacob, 1997.

parvient à défendre vigoureusement son approche originale, qui suscitera très certainement la discussion.

Finalement, Joëlle Proust aborde un problème lié à la question de la nature du sujet d'attitudes propositionnelles. Alors que la question ontologique de la nature des états mentaux et celle de la possibilité de leur réduction psychophysique sont généralement traitées sur un terrain proprement métaphysique, qui fait abstraction de questions épistémologiques, la question de la nature du sujet pensant peut difficilement se poser en faisant abstraction de considérations épistémiques, car la façon dont un sujet a accès à ses états mentaux, en tant qu'états lui appartenant en propre, c'est-à-dire au titre de sujet, semble être constitutive de la nature d'un sujet pourvu de capacités réflexives qui se manifestent dans des auto-attributions d'états mentaux. Ainsi, la particularité de la question de la nature du sujet réflexif est qu'elle ne peut être abordée uniquement sur le terrain métaphysique, mais elle doit se situer à la fois sur un terrain métaphysique et sur un terrain épistémologique. Dans cette perspective générale, le texte de Proust porte sur la question de savoir comment un sujet peut être justifié dans ses auto-attributions d'états perceptifs à partir d'une expérience perceptive. Ce problème général a divers aspects, et Proust se concentre plus spécifiquement sur le problème de l'unicité entre le sujet de l'expérience et celui de l'auto-attribution, à savoir le problème dit « de la transition » — pour reprendre une expression de Christopher Peacocke¹³. Autrement dit, ce problème est celui de savoir en quoi le sujet d'une auto-attribution faite sur la base d'une expérience perceptive est justifié de croire qu'il est bien lui-même le sujet de cette expérience. Dans un premier temps, Proust montre que, contrairement à certains présupposés cartésiens qui suggèrent qu'une telle transition va de soi, ce n'est pas du tout le cas, et que nous sommes effectivement confrontés à un problème important. Après une discussion critique où elle rejette diverses théories de la transition, dont notamment celles de Sydney Shoemaker et de Christopher Peacocke, Proust développe une approche tout à fait originale qui fait valoir l'idée selon laquelle ce serait la nécessité de l'unicité du sujet dans ses pratiques d'actions rationnelles qui fournirait une solution au problème de la transition. Cette thèse est brillamment illustrée à l'aide de divers scénarios possibles dans lesquels un sujet n'aurait pas le plein statut d'agent rationnel, qui est requis pour l'unicité dans la transition de l'expérience perceptive à l'auto-attribution.

13. Voir Peacocke, C., *Being Known*, Oxford, Clarendon Press, 1999.

L'émergence, les modèles de réduction et le mental

Jaegwon Kim

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004937ar>
DOI : <https://doi.org/10.7202/004937ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Kim, J. (2000). L'émergence, les modèles de réduction et le mental.
Philosophiques, 27 (1), 11–26. <https://doi.org/10.7202/004937ar>

Résumé de l'article

Une des doctrines centrales de l'émergentisme est la thèse selon laquelle certaines propriétés d'un tout sont émergentes, en ce sens qu'elles sont irréductibles aux propriétés de base dont elles émergent — c'est-à-dire qu'elles ne peuvent ni être prédites, ni être expliquées à partir de leurs conditions sous-jacentes. Pour comprendre et évaluer cette thèse correctement, il est essentiel que nous disposions d'un concept adéquat de réduction. Nous examinons d'abord le modèle classique de la réduction interthéorique de Nagel, et nous soutenons qu'il ne nous fournit pas une base adéquate pour comprendre la thèse émergentiste. Nous proposons ensuite un autre modèle de réduction, celui de la « réduction fonctionnelle », et nous montrons qu'il constitue une base adéquate permettant d'évaluer la thèse émergentiste. Nous concluons avec une brève discussion de la question de savoir si les propriétés phénoménales d'expériences conscientes, ou *qualia*, sont émergentes.

L'émergence, les modèles de réduction et le mental

JAEGWON KIM

Jaegwon_Kim@brown.edu

Department of Philosophy

Brown University

RÉSUMÉ. — Une des doctrines centrales de l'émergentisme est la thèse selon laquelle certaines propriétés d'un tout sont émergentes, en ce sens qu'elles sont irréductibles aux propriétés de base dont elles émergent — c'est-à-dire qu'elles ne peuvent ni être prédites, ni être expliquées à partir de leurs conditions sous-jacentes. Pour comprendre et évaluer cette thèse correctement, il est essentiel que nous disposions d'un concept adéquat de réduction. Nous examinons d'abord le modèle classique de la réduction interthéorique de Nagel, et nous soutenons qu'il ne nous fournit pas une base adéquate pour comprendre la thèse émergentiste. Nous proposons ensuite un autre modèle de réduction, celui de la « réduction fonctionnelle », et nous montrons qu'il constitue une base adéquate permettant d'évaluer la thèse émergentiste. Nous concluons avec une brève discussion de la question de savoir si les propriétés phénoménales d'expériences conscientes, ou *qualia*, sont émergentes.

ABSTRACT. — A central doctrine of emergentism is the thesis that some properties of wholes are emergent in the sense that they are irreducible to the “ basal ” properties from which they emerge — that is, they are neither predictable nor explainable in terms of their underlying conditions. To understand and properly evaluate this claim, it is essential that we have at hand an appropriate concept of reduction. This paper first examines the classic Nagel model of inter-theoretic reduction and finds it wanting as a basis for understanding the emergentist claim. Another model of reduction, “ functional reduction ”, is proposed, and it is argued that this model does provide an appropriate basis for evaluating the claim that certain properties of wholes are emergent. The paper closes with a brief discussion of the question whether *qualia*, or the phenomenal properties of conscious experience, are emergent.

1. Introduction

L'idée centrale à la notion d'émergence est que, lorsqu'un système composé d'agréats de matière atteint un certain niveau de complexité organisationnelle, il commence à exhiber de nouvelles propriétés jusqu'alors inconnues, des propriétés « émergentes » — propriétés dont l'occurrence n'aurait pu être *prédite* sur la base des propriétés et relations structurales caractérisant les parties constituantes du système. Cette idée s'accompagne d'une autre idée voulant que l'émergence de telles propriétés ne puisse être *expliquée* à partir des processus sous-jacents (« les conditions de base ») desquels elles émergent. Bref, un tout complexe posséderait de nouvelles propriétés qui sont irréductibles aux propriétés et relations de ses parties. De plus, les propriétés émergentes sont conçues comme étant des propriétés authentiques en

ce sens qu'elles posséderaient des pouvoirs causaux distincts qui leur sont propres. Les systèmes complexes peuvent aussi avoir, et ont effectivement, des propriétés non émergentes, des propriétés réductibles aux propriétés de leurs parties, que les émergentistes britanniques, plus tôt au cours de ce siècle, ont appelées des propriétés « résultantes » ou « conséquentielles ».

On fait habituellement remonter l'idée d'émergence à John Stuart Mill et à sa distinction entre propriétés « homopathiques » et « hétéropathiques », mais la terminologie plus familière de propriétés « émergentes » et « résultantes » est due à G. H. Lewes, un contemporain de Mill¹. Cependant, la distinction entre ces deux types de propriétés remonte apparemment à beaucoup plus loin, soit aux Grecs et en particulier à Galien², ce qui montre bien qu'il y a sûrement quelque chose de naturel et d'intuitif dans l'idée d'émergence. Depuis que ce concept a commencé à faire l'objet d'une attention philosophique particulière, comme ce fut le cas au début du xx^e siècle, il a exercé un puissant et persistant attrait sur des penseurs de différents domaines, y compris sur certains scientifiques comme, par exemple, le neurologue réputé Roger W. Sperry. Je crois que c'est cet attrait intuitif du concept qui a permis de le garder bien en vie au milieu du siècle, malgré les influences marquées du positivisme logique et d'autres courants réductionnistes en philosophie et en science. Il fut un temps, pas si lointain, où l'émergentisme était communément associé à des curiosités métaphysiques largement méprisées, comme le néovitalisme et la doctrine de l'élan vital. Mais le climat intellectuel a radicalement changé au cours des quelques trente dernières années, et le vocabulaire de l'émergence semble être maintenant complètement réhabilité. L'utilisation de l'expression « émergence » et de termes de la même famille se retrouve non seulement dans des textes philosophiques mais aussi dans certains écrits purement scientifiques, surtout dans des domaines comme les sciences cognitives, la psychologie, les neurosciences, la théorie de systèmes, la théorie de la complexité, la théorie de l'auto-organisation, etc. Les termes « émerge » et « émergent » semblent venir assez naturellement aux auteurs dans ces divers domaines, ce qui suggère qu'il y a un contenu substantiel partagé par tous, et cela indépendamment d'un engagement philosophique tendancieux. Aussi, il n'est pas étonnant qu'une perspective largement émergentiste ait fait un retour en force dans les discussions philosophiques portant sur la nature du mental et de sa relation aux processus neurologiques sous-jacents.

Considérons la doctrine du physicalisme non réductionniste qui, depuis les trois dernières décennies, constitue l'orthodoxie relativement au problème corps/esprit (*mind/body problem*). Selon ce point de vue, bien que toutes cho-

1. Mill, John Stuart, *A System of Logic*, 8^e édition, London, Longmans, Green, Reader and Dyer, 1872, Livre III, ch. iv, paru d'abord en 1843 ; Lewes, G. H., *Problems of Life and Mind*, London, Kegan Paul, Trench, Tubner & Co., 1875.

2. Voir Caston, Victor, « Epiphenomenalisms, Ancient and Modern », *Philosophical Review*, 106, 1997, p. 309-333.

ses capables de mentalité, comme toutes choses en ce monde, soient constituées exclusivement de particules matérielles et d'agrégats de telles particules, les propriétés psychologiques qu'elles exhibent seraient irréductibles à leurs propriétés physiques et elles formeraient ainsi leur propre domaine. De plus, on pense que ces propriétés psychologiques se comportent en accord avec des lois psychologiques irréductibles, qui permettent de produire des explications causales et des prédictions des phénomènes dans le domaine d'investigation scientifique qui leur est propre. Bref, les propriétés psychologiques sont conçues comme des propriétés causales/nomiques d'ordre supérieur qui émergent de configurations complexes de constituants matériels de niveau inférieur, constituants qui eux-mêmes ne posséderaient aucune propriété psychologique. Cette conception du statut de la psychologie et des autres sciences de niveau supérieur est attrayante à plusieurs égards. D'une part, elle renonce au simple dualisme des substances en reconnaissant que les substrats physiques sont nécessaires et suffisants aux propriétés cognitives/psychologiques de niveau supérieur. Et, d'autre part, elle assure à la psychologie une indépendance et une intégrité disciplinaire comme science autonome, sans pour cela devoir accepter l'existence d'entités métaphysiques douteuses, comme les esprits immatériels, les entéléchies ou autres choses semblables. De plus, cette conception peut être généralisée à d'autres sciences spéciales : selon ce point de vue, chaque science spéciale a son propre sujet distinct, constitué par un ensemble de propriétés d'ordre supérieur qui sont irréductibles aux sciences sous-jacentes, et en particulier à la physique élémentaire ; chacune des sciences spéciales cherche des lois causales et tente de formuler des explications fondées sur des lois à l'intérieur de son propre domaine.

La ressemblance entre l'émergentisme et le physicalisme non réductionniste est évidente. Les premiers émergentistes, comme Samuel Alexander, C. Lloyd Morgan et C. D. Broad, auraient trouvé tout cela assez familier et parfaitement naturel, comme un vieil ami perdu et réapparu apparemment inchangé par des épreuves difficiles affrontées au cours des ans.

2. Réduction nagelienne et lois de correspondance

Pour évaluer la situation actuelle impliquant l'émergentisme et le physicalisme non réductionniste, nous devons comprendre clairement le concept de réduction puisqu'il est évidemment au cœur de ces deux doctrines. Comme nous l'avons vu, les propriétés émergentes sont ces propriétés de systèmes complexes qui ne sont pas réductibles aux propriétés et aux relations de leurs constituants — c'est-à-dire que ces propriétés ne sont ni prédictibles ni explicables par des propriétés ou des relations de niveau inférieur. Pour comprendre ce qu'est une propriété émergente nous devons donc comprendre ce qui rend une propriété réductible ou irréductible. Et pour comprendre cela, nous devons comprendre ce qu'est une réduction — c'est-à-dire que nous devons disposer d'un modèle adéquat de la réduction. Comprendre la réduction est

également crucial pour comprendre le physicalisme non réductionniste. Comme nous l'avons vu, la thèse centrale de cette doctrine est essentiellement la même que celle de l'émergentisme : tous les organismes, comme les autres objets de ce monde, sont entièrement faits de fragments de matière, mais certains exhibent certaines propriétés, parmi lesquelles des propriétés mentales, qui ne sont pas réductibles aux propriétés de ces fragments de matière. Mais en fait, comme nous le verrons, ces deux doctrines divergent de façon substantielle, et cela malgré leurs ressemblances superficielles. Plus précisément, nous verrons que ces deux positions sont fondées sur deux notions de réduction passablement différentes et que, par conséquent, du point de vue d'une conception émergentiste de la réduction, certaines versions influentes du physicalisme non réductionniste devraient être considérées comme des théories réductionnistes.

Le consensus antiréductionniste — pour reprendre l'expression bien choisie de Ned Block³ — qui a régné depuis la fin des années 1960 s'appuie sur une compréhension de la notion de réduction telle qu'elle est entendue dans le modèle classique de la réduction interthéorique développé par Ernest Nagel, dans les années 1950 et 1960⁴. Selon Nagel, réduire une théorie T , comprenant les prédicats primitifs $F_1, \dots, F_n$, à une théorie de base T^* , comprenant un ensemble disjoint de prédicats $G_1, \dots, G_n$, c'est dériver chaque loi de T à partir des lois de T^* à l'aide de lois de correspondance (*bridge laws*) liant chaque prédicat primitif de T à un prédicat, vraisemblablement complexe, de T^* . Ces lois de correspondance ont la forme suivante :

Pour tout $x_1, \dots, x_n$, $F_i(x_1, \dots, x_n)$ si, et seulement si, $G(x_1, \dots, x_n)$

où G est un prédicat complexe composé de $G_1, \dots, G_m$. Ces lois de correspondance, liant les prédicats de la théorie cible à ceux de la théorie de base, sont au cœur de la réduction nagelienne. Cela est évident du fait qu'elles sont non seulement *nécessaires* pour établir des liens déductifs entre les deux théories, formulées dans des vocabulaires hétérogènes, mais elles sont aussi *suffisantes* pour la réduction nagelienne en ce sens que lorsqu'elles sont en place, cela assure une dérivation réductrice des lois- T à partir des lois- T^* *quels que soient les contenus spécifiques de T et de T^** ⁵. Cet aspect de la réduction est important, car il rend possible la discussion des possibilités de réduction avant même que les théories pertinentes n'aient été *complètement*

3. Voir Block, Ned, « Anti-Reductionism Slaps Back », *Philosophical Perspectives*, 11, 1997, p. 107-132.

4. Nagel, Ernest, *The Structure of Science*, New York, Harcourt, Brace and World, 1961, ch. 11.

5. La raison pour laquelle les lois de correspondance garantissent la réduction nagelienne est qu'elles permettent la « traduction » des lois de la théorie à réduire par des énoncés formulés exclusivement dans le vocabulaire de la théorie réductrice, c'est-à-dire par des lois de la théorie réductrice. Pour plus de détails, voir Kim, Jaegwon, « Making Sense of Emergence », *Philosophical Studies*, 95, 1999, p. 3-36.

formulées car, bien évidemment, il n'est pas réaliste d'espérer atteindre un tel degré d'achèvement, quelles que soient les théories considérées, dans un domaine assez riche. Bien sûr, nous devons tout de même savoir quelque chose des prédicats des théories en cause. Nous devons d'une part comprendre ces prédicats tels qu'ils sont actuellement utilisés dans les théories, mais nous devons également comprendre comment le vocabulaire de chaque théorie pourrait éventuellement être étendu. Autrement dit, nous avons besoin d'une idée de ce qui fait qu'un prédicat est un prédicat de T ou un prédicat de T^* . Par exemple, nous voulons et devons être en mesure de discuter la réductibilité de la psychologie cognitive à la neurobiologie avant de connaître la forme achevée de ces deux théories, car un tel achèvement risque de n'être jamais atteint de notre vivant, et il risque même de ne jamais être atteint tout à fait. Cependant, nous sommes dans une meilleure position vis-à-vis des prédicats. Bien que nos concepts et les prédicats qui les expriment puissent changer selon les vicissitudes de nos théories, ils sont sans aucun doute plus stables que les théories. Après tout, des théories en compétition peuvent être formulées dans le même vocabulaire et, dans la plupart des cas, nous sommes capables de dire quand un prédicat est un prédicat psychologique, un prédicat biologique ou quelque chose d'autre. Et il est possible de construire des arguments réductionnistes ou antiréductionnistes simplement sur la base des caractères généraux des prédicats psychologiques et physiques, indépendamment d'exemples spécifiques.

Il est donc tout à fait compréhensible que le principal argument anti-réductionniste, qui fut en grande partie responsable du déclin rapide du réductionnisme au cours des années 1970, ait porté essentiellement sur la disponibilité de lois de correspondance liant les prédicats psychologiques aux prédicats physiques (ou les propriétés psychologiques aux propriétés physiques, pour parler en termes métaphysiques). Il s'agit, bien sûr, du célèbre argument de la réalisation multiple qui, comme nous le savons, a eu une très grande influence⁶. Cet argument ne concernait pas — ou à tout le moins pas directement — la question de savoir si les lois physiques ou biologiques pourraient suffire à produire des lois psychologiques de façon purement déductive ; l'argument était totalement indépendant des contenus particuliers des théories psychologiques et physiques, quels qu'ils aient pu être. Mais il mettait plutôt l'accent sur la question suivante : *Existe-t-il des lois de correspondance psychophysiques*? Et, comme nous le savons, l'argument de la réalisation multiple répond négativement à cette question. Chaque propriété psychologique a un nombre indéfini de réalisateurs divers au niveau physique, et il est impossible de déterminer, à l'aide d'une loi de correspondance ayant la forme notée plus haut, quelle propriété physique serait le corrélat

6. Selon moi, cet argument a également été très mal compris. À ce sujet, voir Kim, J., « Multiple Realization and the Metaphysics of Reduction », dans Kim, J., *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993.

d'une propriété psychologique⁷. Cet argument est habituellement complété par un argument additionnel montrant qu'il serait inadéquat de formuler une loi de correspondance en utilisant, comme corrélât physique, une disjonction de tous les réalisateurs possibles d'une propriété mentale⁸.

Ces considérations conduisent assez naturellement à une conception des propriétés psychologiques comme étant des propriétés formelles abstraites qui sont indépendantes de leur composition matérielle ou du mécanisme responsable de leur implémentation. Elles conduisent également à une certaine conception de la nature de la psychologie comme science. Brièvement, l'idée est que les propriétés psychologiques font abstraction des détails physiques/compositionnels concrets de ce qui les réalise, et que l'étude scientifique de ces propriétés, des lois qui les gouvernent et des explications des phénomènes psychologiques peut être menée sans tenir compte des implémentations ou mécanismes physiques qui les sous-tendent. C'est en ce sens que la psychologie et les autres sciences « spéciales » ont été dites autonomes relativement aux sciences physiques/biologiques sous-jacentes.

En résumé, on peut dire que le principal argument contre la réduction psychophysique, qui a engendré les formes de physicalisme non réductionniste présentement populaires, est l'argument de la réalisation multiple, et que celui-ci repose essentiellement sur l'affirmation qu'il n'existe pas de lois de correspondance psychophysiques, ce qui rendrait impossible une dérivation nagelienne de lois psychologiques à partir de lois physiques. Toutefois, ce que l'on note plus rarement ce sont les présuppositions et les implications physicalistes/réductionnistes de cet argument. En effet, dans cet argument, on retrouve implicitement la proposition que les propriétés mentales sont physiquement réalisées, et même la proposition plus forte que chaque fois qu'une propriété mentale est exemplifiée dans un organisme (ou un système), ce serait en vertu du fait que certaines propriétés physiques de l'organisme la réalisent, ou fournissent un mécanisme l'implémentant dans cet organisme. En ce sens, n'importe quelle forme de non-réductionnisme qui s'inspire de l'argument de la réalisation multiple est en fait une doctrine physicaliste de façon non triviale. Et, comme nous le verrons, une telle approche mène ultimement au réductionnisme, une position diamétralement opposée à celle de l'émergentisme. Selon mon point de vue, le fonctionnalisme, tel qu'il est aujourd'hui largement répandu, est un exemple d'une telle position. Cette façon de traiter le problème de la relation corps/esprit a été communément acceptée comme étant une approche non réductionniste⁹, mais comme nous

7. C'est Jerry Fodor qui a proposé la version la plus influente de cet argument. Voir Fodor, J. A., « Special Sciences », réédité dans Fodor, J. A., *Representations*, Cambridge, Cambridge University Press, 1981. (Cet article a paru initialement en 1974.)

8. Sur cette question, le lecteur intéressé pourra consulter les articles de Kim et de Fodor mentionnés aux notes 6 et 7.

9. Ce ne sont pas tous les fonctionnalistes qui ont endossé cette thèse. David Armstrong et David Lewis sont parmi les rares d'entre eux qui ne l'ont pas fait. L'attitude des fonctionnalistes

le verrons dans ce qui suit, dès lors que nous adoptons un modèle de la réduction plus adéquat que le modèle nagelien, il apparaît clairement que le fonctionnalisme est en fait une forme de réductionnisme.

3. Un *desideratum* de réduction, et la réduction fonctionnelle

La psychologie est-elle réductible, au sens de la réduction nagelienne, à la neurobiologie ou à une autre théorie non psychologique sous-jacente? Avant de tenter de répondre à cette question, nous devons d'abord répondre à une autre question concernant la signification philosophique d'une réduction nagelienne de la psychologie. Si nous supposons que la psychologie a effectivement été réduite en ce sens, qu'est-ce que cela montrerait? La réponse est que cela montrerait bien peu de choses, car selon les théories dualistes classiques concernant la relation corps/esprit — comme par exemple la théorie spinoziste du double aspect, la doctrine de l'harmonie préétablie de Leibniz, le parallélisme ou le monisme neutre — les lois psychophysiques seraient disponibles en abondance. En fait, l'existence de telles lois n'est même pas exclue par le dualisme cartésien, bien qu'elle soit probablement exclue par la théorie spécifique que Descartes a effectivement soutenue — du moins pour ce qui concerne certains états dotés de contenu (soit les « pensées rationnelles »). Cela signifie que la réductibilité nagelienne du mental est non seulement compatible avec ces théories dualistes, mais qu'elle est en fait impliquée par plusieurs d'entre elles. Si la réduction psychophysique doit être comprise sur la base du modèle de la réduction nagelienne, le réductionnisme n'aurait donc pas de conséquences métaphysiques intéressantes. Mais alors, pourquoi devrait-on se préoccuper de la réduction de l'esprit au corps ou de réductionnisme?

Cela jette un doute sur l'adéquation de la réduction nagelienne comme base de discussion du problème corps/esprit, et sape la pertinence philosophique des arguments antiréductionnistes qui reposent sur ce modèle.

Considérons la réduction nagelienne à partir du point de vue émergentiste. La principale question que les émergentistes avaient à l'esprit, à propos de la relation entre les propriétés de niveau supérieur et leurs conditions de base, était la suivante : l'occurrence d'une propriété de niveau supérieur peut-elle être prédite *uniquement* à partir d'informations concernant le niveau de base? La question concernant les phénomènes mentaux serait donc la suivante : l'occurrence de la douleur, par exemple, peut-elle être prédite *uniquement* à partir de la connaissance de la physiologie du cerveau? Cette première question s'accompagne également d'une question explicative : peut-on expliquer l'occurrence d'une propriété émergente *uniquement* sur la base

envers le réductionnisme a été marquée par une ambivalence profonde et une instabilité incertaine, et cela même si ce sont plutôt les vues de certains antiréductionnistes qui ont prédominé. J'espère que le présent article permettra d'expliquer et de résoudre cette ambivalence.

d'une connaissance du niveau de base? Peut-on expliquer pourquoi vous ressentez de la douleur par exemple, plutôt qu'un picotement, uniquement sur la base de faits physiques/biologiques à propos de vous? Si une réduction d'une théorie d'un domaine M à une théorie de base ayant pour domaine P ne fournit pas de prédictions ou d'explications d'une occurrence des phénomènes de M sur la base de faits dans P , que peut donc nous procurer une telle réduction?

Ces considérations nous amènent à la suggestion suivante : si la notion de réduction doit avoir un rôle significatif dans les débats sur la relation corps/esprit, alors elle doit satisfaire le *desideratum* suivant :

Si une propriété d'ordre supérieur P est réductible à un niveau inférieur L , alors l'occurrence de P doit être prédictible et explicable uniquement sur la base d'informations concernant les faits de niveau L .

Si nous avons un modèle de la réduction satisfaisant cette exigence, cela pourrait motiver une définition des propriétés émergentes comme étant ces propriétés de niveau supérieur qui ne sont pas réductibles, selon ce modèle, à des faits d'un niveau inférieur quelconque¹⁰. Une telle compréhension de la notion d'émergence a pour conséquences heureuses qu'une propriété est émergente relativement à ses conditions de base seulement si elle ne leur est pas réductible, et que l'émergentisme est vrai d'un ensemble de propriétés seulement si le réductionnisme est faux de ces propriétés.

Selon les émergentistes, il existe bien sûr certaines propriétés d'un tout qui *sont* effectivement prédictibles sur la base des propriétés de leurs parties. Ils citeraient en exemple la masse : la masse de cette table peut être dérivée, logiquement ou mathématiquement, de la masse de chacune de ses parties, disons le dessus de la table et sa structure¹¹. De plus, nous pouvons expliquer pourquoi la table a cette masse à partir d'informations à propos de la masse de ses parties. Mais, toujours selon les émergentistes, diverses propriétés de composés chimiques, comme par exemple la transparence et la viscosité de l'eau, ne seraient pas dérivables, et donc non prédictibles à partir des propriétés de leurs atomes constitutants ; c'est-à-dire que les configurations structurales au niveau des molécules d'eau ne seraient pas dérivables à partir des propriétés des atomes d'hydrogène et d'oxygène considérées isolément. Cette affirmation des émergentistes paraît plausible, du moins à première vue, en raison de l'absence des concepts de transparence et de viscosité dans la physique atomique. Et, toujours selon ce point de vue, cela vaudrait également pour la mentalité et la conscience, ce qui veut dire qu'on peut tout savoir des processus physiques et biologiques se déroulant dans un organisme sans pour

10. Suivant une pratique répandue, nous utilisons ces notions impliquant différents « niveaux » en un sens intuitif — particulièrement les notions de « supérieur » et « inférieur » telles qu'appliquées aux niveaux. Ces notions sont discutées de façon plus détaillée dans Kim, J., « The Metaphysics of a Layered World », à paraître.

11. Il est douteux cependant que cette thèse émergentiste soit vraie. L'additivité de la masse serait aujourd'hui considérée comme une propriété empirique et contingente.

autant connaître, de ce fait, dans quel état de conscience il se trouve, si état de conscience il y a — c'est-à-dire sans savoir s'il a ou non une expérience consciente et, si c'est le cas, quel en serait le caractère qualitatif.

Un trait intéressant du *desideratum* sur la réduction, que nous avons formulé plus haut, est qu'il n'exige pas que la propriété *M*, visée par la réduction, soit identifiée à une propriété ou même à un groupe ou à une disjonction de propriétés au niveau de base. (En fait, il n'exige même pas qu'il y ait des lois de correspondance nagelienne.) L'explication ou la prédiction d'une exemplification de *M*, en une occasion particulière, peut être « locale » au lieu de viser à être « globale » en ce sens qu'elle peut recevoir une explication réductrice qui en appelle à un mécanisme ou à une structure sous-jacente impliquée dans *M* en cette occasion particulière, alors qu'une autre exemplification de *M* pourrait recevoir une explication invoquant un mécanisme complètement différent. La raison en est que rien ne garantit qu'un seul et unique mécanisme ou une seule et unique structure soient sous-jacents à toutes les exemplifications actuelles et possibles de *M* ; en fait, il est plausible que le contraire soit la norme. Ce qui, bien entendu, n'est qu'une autre facette du phénomène de la réalisation multiple. Cela signifie qu'un modèle de réduction peut satisfaire le *desideratum* sans exiger de réductions type/type, et donc que les modèles qui satisfont ce *desideratum* peuvent le faire tout en conciliant la réalisation multiple des propriétés réduites.

Je vais maintenant esquisser un modèle de réduction qui satisfait le *desideratum* formulé plus haut. C'est ce que j'ai appelé la « réduction fonctionnelle », qui consiste dans les trois étapes suivantes :

Étape 1. Fonctionnaliser la propriété visée par la réduction — c'est-à-dire donner une caractérisation de la propriété par son rôle causal qui est spécifié dans les termes des propriétés au niveau de base.

Une telle caractérisation est appelée une *définition fonctionnelle* de la propriété en question. La première étape dans la réduction du gène est de le comprendre fonctionnellement, c'est-à-dire dans les termes du « rôle causal » qu'il est supposé remplir. Ainsi, on explique le gène comme ce mécanisme qui, dans un organisme, accomplit la tâche d'encoder et de transmettre l'information génétique.

Étape 2. Identifier la ou les propriétés (ou le ou les mécanismes) présents au niveau de base, qui accomplissent les tâches causales spécifiées dans le ou les systèmes examinés — c'est-à-dire trouver le réalisateur de la propriété fonctionnalisée pour les systèmes examinés.

Au contraire de l'étape 1, l'étape 2 est carrément dans le domaine de la recherche empirique/théorique. Les philosophes n'ont aucune contribution à apporter à cette étape. Notons que, étant donné le phénomène de réalisation multiple, l'étape 2 sera normalement exécutée cas par cas, car il arrivera rarement que la recherche scientifique vise et réussisse à découvrir tous les réalisateurs nomologiquement possibles implémentant, à un niveau inférieur, la

propriété ciblée — ou qu'il y ait un unique mécanisme sous-jacent qui puisse être mis au jour. Il est plus plausible que les chercheurs tenteront d'identifier le ou les réalisateurs impliqués dans une espèce particulière ou dans un groupe de structures qui présentent un intérêt pour leur programme de recherche. Ainsi, pour les organismes terrestres, les molécules d'ADN furent identifiées comme porteurs et transmetteurs de l'information génétique. Mais il est nomologiquement possible que différentes molécules fassent ce travail biologique dans différentes espèces d'organismes terrestres. Qui plus est, dans des mondes possibles où prévaudraient des lois physiques différentes des nôtres, des molécules autres que celles d'ADN pourraient accomplir cette tâche.

Étape 3. Développer une théorie qui explique comment les réalisateurs de la propriété cible font pour accomplir la tâche causale spécifiée.

Nous voulons une théorie qui explique comment les molécules d'ADN font pour encoder et transmettre l'information génétique des parents à leur progéniture. Une compréhension réductrice du gène — c'est-à-dire une compréhension du gène en termes de processus sous-jacents — devrait, semble-t-il, inclure une telle théorie. De toute façon, les étapes 2 et 3 seront presque certainement accomplies en collaboration : on peut s'attendre à ce qu'elles constituent deux parties, ou deux aspects, d'un même programme de recherche. Il est peu probable que la recherche puisse identifier les réalisateurs d'une propriété scientifiquement significative sans avoir recours à une théorie explicative des phénomènes sous-jacents mis en jeu. Au-delà de la caractérisation causale donnée par la définition fonctionnelle de la propriété cible, une théorie explicative en développement aidera à circonscrire et à désigner les endroits où les réalisateurs pourraient se trouver¹².

4. La satisfaction du *desideratum* de réduction

Nous allons maintenant voir comment notre *desideratum* sur la réduction — le réquisit émergentiste d'explication et de prédiction — peut être satisfait si

12. Ces idées proviennent, à l'origine, des certains écrits de David Armstrong et de David Lewis portant sur une version du matérialisme qui, à l'époque, était connue sous l'appellation de « Central State Materialism », bien qu'ils ne les aient pas proposées explicitement comme fournissant un modèle de la réduction. Voir Armstrong, D., *A Materialist Theory of Mind*, London, Routledge & Kegan Paul, 1968 ; ainsi que Lewis, D., « An Argument for the Identity Theory », *Journal of Philosophy*, 67, 1970, 203-211. Certaines idées similaires ont récemment refait surface, étant cette fois plus clairement identifiées à la réduction et à l'explication réductrice chez Joseph Levine dans son article « On Leaving Out What It's Like », dans Davies, M. et Humphreys, G., dir., *Consciousness*, Oxford, Blackwell, 1993. Ce que je propose ici est que ce modèle, et non pas le modèle de Nagel, constitue un modèle adéquat de la réduction, et cela non seulement pour ce qui concerne le débat sur le problème corps/esprit, mais pour la réduction en général. Pour une discussion plus élaborée, voir Kim, J., *Mind in a Physical World*, Cambridge, MIT Press, 1998 ; et « Making Sense of Emergence », *Philosophical Studies*, 95, 1999, p. 3-36.

une propriété est réductible au sens du modèle de la réduction fonctionnelle que nous venons d'esquisser.

Considérons d'abord la prédiction de l'occurrence d'une propriété *M* fonctionnellement réduite : est-ce qu'un système donné *S* exemplifie *M* au temps *t*? Pour répondre affirmativement à cette question, en accord avec la caractérisation fonctionnelle de *S*, *S* devrait exemplifier, à *t*, une propriété *P* remplissant un certain rôle causal *R*. Comme nous le supposons dans ce cas, une réduction fonctionnelle a été accomplie, et nous supposons donc comme étant acquise la connaissance du réalisateur de la propriété *M* dans des systèmes comme *S*, à savoir *P* qui est une propriété ou un mécanisme du niveau de base. La question prédictive est donc la suivante : est-ce que *S* exemplifie *P* au temps *t*? Manifestement, il s'agit là d'une question qui peut recevoir une réponse au niveau de base, ce qui signifie que les occurrences de *M* peuvent effectivement être prédites à partir d'informations concernant uniquement le niveau de base. (Il peut y avoir plus d'un réalisateur de *M* dans des systèmes comme *S*, mais cela ne fait aucune différence.)

Tournons-nous maintenant vers l'explication de l'occurrence d'une propriété *M*. Pourquoi le système *S* exemplifie-t-il *M* au temps *t*? Encore une fois, pour que *S* exemplifie *M* à *t*, *S* doit exemplifier un réalisateur de *M*, à *t*. *S* exemplifie donc *P* à *t*, et *P* est un réalisateur de *M* — *P* satisfait le rôle causal spécifié pour *M*. Et c'est pourquoi *S* exemplifie *M* à *t*.

Considérons l'exemple suivant : Jean éprouvera-t-il de la douleur au temps *t*? Éprouver de la douleur est pour lui, par définition, être dans un état (c'est-à-dire exemplifier une propriété) susceptible de causer des cris et des hurlements, et un comportement de fuite. Chez les humains, l'activation de fibres *C* est le réalisateur de la douleur. La question est donc de savoir si les fibres *C* de Jean seront activées au temps *t*. Et ce sont des faits et des lois au niveau physique/biologique qui permettront, en principe, de répondre à cette question.

Pourquoi Jean éprouve-t-il de la douleur au temps *t* (au lieu d'un picotement, par exemple)? Avoir de la douleur c'est se trouver dans un état susceptible de causer des cris et des hurlements, et un comportement de fuite. Au temps *t*, les fibres *C* de Jean sont activées et l'activation des fibres *C* est un réalisateur de la douleur chez les humains. (Et la stimulation des fibres *C* n'est pas le réalisateur de picotements chez les humains.) Voilà pourquoi Jean éprouve de la douleur, et non pas un picotement, au temps *t*.

Il me semble que, dans ces deux cas, nous sommes en présence d'une véritable prédiction et d'une véritable explication. De plus, celles-ci sont *réductrices* en ce sens que les ressources explicatives et prédictives qui sont mises à profit proviennent exclusivement du niveau réducteur de base. C'est une définition qui nous fournit les relations réductrices entre le niveau de base et la propriété cible, en l'occurrence une caractérisation fonctionnelle de *M* en termes de relations causales impliquant des éléments du niveau de base.

Comparons cela à la situation impliquant une réduction nagelienne de *M*. Supposons qu'une théorie à laquelle *M* appartient ait été réduite à une théorie de niveau de base à l'aide de lois de correspondance incluant celle-ci :

M T

où *T* est une propriété de niveau de base. Étant donné une telle réduction, comment prédire ou expliquer, de façon réductrice, une occurrence de *M*?

Notons que la propriété *M* n'a pas d'occurrence dans la théorie de base initiale, et que c'est seulement la théorie de base enrichie — à savoir la théorie initiale plus les lois de correspondance — qui inclut *M*. En fait, c'est par le biais de la loi de correspondance notée plus haut que *M* apparaît dans la théorie enrichie. Ainsi, la seule façon de prédire des occurrences de *M* serait d'utiliser cette loi de correspondance, essentiellement de la manière suivante :

Le système *S* exemplifiera *T* à *t*

M T

Donc, *S* exemplifiera *M* à *t*.

Notons cependant que cette prédiction ne satisfait pas le *desideratum* sur la réduction. Car, bien que ce soit une prédiction inductive parfaitement acceptable, elle s'établit sur une base comprenant la loi de correspondance. Et celle-ci véhicule de l'information préalable au sujet de la corrélation entre les occurrences de *M* et celles de *T*, ce qui constitue une connaissance présupposant des observations préalables de *M*. De telles prédictions d'occurrences de *M* ne sont donc pas fondées *uniquement* sur la base d'informations concernant le phénomène et les lois au niveau de base et, par conséquent, elles ne satisfont pas le *desideratum* sur la réduction.

Les émergentistes savaient bien que nous faisons couramment des prédictions inductives d'états conscients sur la base de corrélations observées entre ces états et des états comportementaux ou neuronaux. Mais la question qui avait pour eux un intérêt fondamental était plutôt la suivante : étant donné tout ce que l'on peut savoir au sujet du comportement et du système neuronal d'un organisme, et étant donné cette *seule* information, est-il possible de prédire ou d'inférer que cet organisme éprouve de la douleur ou qu'il est conscient? Leur réponse était que cela est impossible. Il ne suffirait pas de répliquer que l'on peut inférer que l'organisme doit éprouver de la douleur parce qu'il a un état cérébral similaire à celui des humains lorsque ceux-ci éprouvent de la douleur. Car une telle réponse fait encore une fois appel à une connaissance préalable des corrélations entre la douleur et les états du cerveau.

Nous pouvons donc conclure que les réductions se conformant au modèle de Nagel, contrairement aux réductions fonctionnelles, ne fournissent ni des prédictions réductrices ni des explications réductrices. Elles ne satisfont tout simplement pas le *desideratum* sur les réductions que nous avons formulé plus haut.

5. La conscience est-elle réductible ou émergente?

Revenons maintenant à la réduction de l'esprit. Est-ce que la mentalité — c'est-à-dire l'ensemble des diverses propriétés mentales — peut être réduite? Et si elle le peut, quelle serait sa base de réduction? Le cerveau et le comportement sont bien sûr deux candidats assez évidents à ce titre. Le behaviorisme logique ou analytique fut une tentative pour réduire les propriétés mentales en les définissant en termes de comportements ou de dispositions comportementales. Mais il est largement admis que cette tentative a échoué. Comme on l'a fréquemment souligné, les concepts mentaux ou les propriétés mentales ne peuvent tout simplement pas être définis en termes de dispositions comportementales.

La théorie de l'identité type/type de Smart et Feigl¹³, qui a brièvement régné à la fin des années 1950 et au début des années 1960, peut être considérée comme une théorie réductionniste modelée sur la conception nage-lienne de la réduction, mais à une différence importante près. Leur suggestion était de remplacer les *bi-conditionnels* psychophysiques constituant les lois de correspondance, qui formulent des corrélations uniformes entre les deux domaines séparés du mental et du physique, par des *identités* psychophysiques. Ainsi, la loi de correspondance corrélationnelle « La douleur se produit chez un organisme au temps *t* si, et seulement si, il a un état cérébral *X* au temps *t* » serait remplacée par une loi d'identité, jouant le rôle du principe de correspondance, à savoir : « La douleur = l'état cérébral *X* », et les autres lois de correspondance seraient remplacées de façon similaire. Mais ces identités, connues sous le nom d'« identités théoriques »¹⁴, étaient considérées comme étant empiriques et contingentes. Or il est clair que toute réduction reposant sur l'intermédiaire de lois de correspondance empiriques, qu'il s'agisse de corrélations ou d'identités, ne parvient pas à satisfaire le *desideratum* sur la réduction. Ce point est particulièrement important dans la mesure où une théorie de l'identité type/type, à la Smart/Feigl, a récemment manifesté les signes d'une renaissance¹⁵. Les défenseurs de cette nouvelle version de la thèse de l'identité type/type caractérisent habituellement les identités de correspondance comme étant empiriques mais nécessaires. Elles sont entendues comme étant des identités nécessaires *a posteriori* dans

13. Voir Feigl, H., « The "Mental" and the "Physical" », *Minnesota Studies in the Philosophy of Science*, 2, Minneapolis, University of Minnesota Press, 1958, p. 370-497 ; et Smart, J. J. C., « Sensations and Brain Processes », *Philosophical Review*, 68, 1959, p. 141-156.

14. Je crois que Hilary Putnam fut le premier à utiliser cette expression (ou plus précisément l'expression « *theoretical identification* ») ; voir Putnam, H., « Minds and Machines », dans Hook, S., dir., *Dimensions of Mind*, New York, New York University Press, 1960.

15. Voir notamment Hill, C. et McLaughlin, B., « There are Fewer Things in Reality than Dreamt of in Chalmers' Philosophy », *Philosophy and Phenomenological Research*, 59, 1999, p. 445-454 ; Block, N. et Stalnaker, R., « Conceptual Analysis, Dualism, and the Explanatory Gap », à paraître dans *Philosophical Review*.

le sens de la thèse influente de Saul Kripke, et cela au même titre par exemple que « l'eau = H_2O » et « la chaleur = le mouvement moléculaire ».

Nous ne pouvons pas nous engager ici dans une discussion élaborée de la question de savoir si, et comment, ce modèle de réduction cerveau/esprit permet de rendre compte de l'explication réductrice¹⁶. Cependant, il est clair que cette forme de réduction ne fournit pas de prédictions du type de celles recherchées par les émergentistes, prédictions codifiées dans notre *desideratum* sur la réduction. Comme ces identités, qu'elles soient nécessaires ou contingentes, ne peuvent être connues qu'empiriquement, et comme elles sont requises pour toute prédiction d'occurrences mentales reposant sur de l'information relative aux neurones, par exemple pour la prédiction d'occurrences de la douleur, il ne sera pas possible de prédire de telles occurrences *uniquement* sur la base d'information neurobiologique ou behaviorale à propos des organismes. Une connaissance préalable de la corrélation entre les états de douleur et ceux du cerveau doit être supposée acquise dans toute prédiction de ce type, ce qui viole le *desideratum* sur la réduction.

Lorsqu'il est appliqué au problème de la relation corps/esprit, le modèle de la réduction fonctionnelle offre l'attrait de rendre intelligible la façon dont le mental est lié aussi bien au comportement qu'au cerveau. Selon ce modèle, le mental dépend conceptuellement du comportement dans la mesure où chaque propriété mentale peut être définie comme étant un intermédiaire causal entre les stimuli, comme inputs, et le comportement (et peut-être d'autres propriétés mentales), comme outputs¹⁷. Cette dépendance est donc conceptuelle et nécessaire. D'autre part, les réalisateurs de ces propriétés mentales fonctionnalisées sont des propriétés et des fonctions neuronales. Le fait qu'une propriété neuronale donnée réalise la douleur, par exemple, est quelque chose d'empirique et de contingent. Empirique pour la raison évidente que la recherche scientifique sophistiquée est requise si nous voulons identifier l'état neuronal qui joue le rôle causal assigné à la douleur ; et contingent puisque le fait qu'une propriété neuronale donnée puisse ou non jouer ce rôle causal dépend, tout comme le caractère intrinsèque de la propriété elle-même, des lois causales qui prévalent dans un monde donné, et nous supposons généralement que ces lois prévalent seulement de façon contingente. Somme toute, selon le modèle de la réduction fonctionnelle, les esprits dépendent du comportement de façon conceptuelle, et du cerveau de façon contingente.

Mais rien de tout cela n'implique que le mental soit en fait fonctionnellement réductible ou non. Selon le modèle de la réduction que je défends ici, cette question revient à la suivante : les propriétés mentales peuvent-elles

16. Cette question est discutée dans Kim, J., « Reduction, Reductive Explanation, and the Explanatory Gap », à paraître.

17. Ou de façon holiste via la ramseyfication ; voir Lewis, D., « Psychophysical and Theoretical Identifications », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

être fonctionnalisées? Là est la question cruciale. La raison en est qu'à partir du moment où nous acceptons le postulat physicaliste selon lequel les propriétés mentales n'ont pas de réalisateurs non physiques, une propriété mentale fonctionnalisée doit avoir des réalisateurs physiques, et ceux-ci devraient pouvoir être découverts, en principe, grâce à la recherche scientifique. Et, puisqu'une propriété mentale est exemplifiée en une occasion particulière, nous savons qu'un de ses réalisateurs physiques doit être réalisé à cette occasion. De sorte que si une propriété mentale a été fonctionnalisée, le reste de la tâche revient à la science, et d'un point de vue métaphysique, c'est comme si elle avait été réduite. Si nous savons qu'une propriété mentale est en principe fonctionnalisable, nous savons qu'elle est réductible.

Cela signifie que le fonctionnalisme classique du type de celui qui a été formulé par Armstrong et Lewis est correct. Le mental serait physiquement réductible, puisqu'il serait fonctionnalisable. Si c'est le cas, est-il réductible au cerveau, comme l'ont préconisé Armstrong et Lewis? Cela dépend : pour ce qui est des propriétés mentales des humains il est pratiquement certain que leurs réalisateurs sont des propriétés du cerveau. Et la même chose doit être vraie de toutes les créatures terrestres douées de mentalité, c'est-à-dire ayant des propriétés mentales. Mais ce n'est qu'un début. Comme nous l'avons vu, la réduction devra procéder cas par cas, et elle procédera vraisemblablement en se concentrant sur l'identification des réalisateurs neuronaux du mental chez les humains et chez des espèces animales proches de l'espèce humaine (ce qui, comme j'en suis sûr, est ce que la recherche scientifique sur la conscience fait présentement). Il ne sera pas possible au cours de l'existence de l'espèce humaine d'identifier tous les réalisateurs actuels et possibles du mental, et en ce sens la réduction corps/esprit ne peut jamais être complétée. Nous n'avons peut-être même pas un intérêt sérieux, à long terme, à aller beaucoup plus loin que les êtres humains et quelques espèces animales qui nous offrent un intérêt particulier. Si nous réussissons à identifier, chez les humains, les réalisateurs neuronaux/physiques de toutes les propriétés mentales que les humains sont capables d'exemplifier, nous aurons alors acquis une compréhension réductrice complète de la mentalité humaine. Rien de plus n'est requis. Il me semble que nos succès dans l'identification, même approximative et sommaire, des réalisateurs neuronaux de certaines propriétés et fonctions sélectionnées chez les humains, suffiraient à faire valoir le réductionnisme corps/esprit, dès lors qu'ils sont combinés à la thèse métaphysique générale du physicalisme selon laquelle les propriétés mentales n'ont de réalisateurs que physiques.

Mais ces commentaires ne s'appliquent qu'aux seules propriétés mentales qui sont fonctionnalisables. Notre dernière question est donc la suivante : est-ce que toutes les propriétés mentales sont fonctionnalisables? Comme nous le savons, cette question a suscité de vives controverses, et nous ne pouvons ici nous engager dans une revue et une discussion des divers problèmes qu'elle soulève. Je me contenterai de signaler mon vote négatif, et cela

sans arguments à l'appui. Il y a en effet de bonnes raisons de croire que les propriétés mentales intentionnelles, comme la croyance et le désir, sont effectivement fonctionnalisables, mais que les caractères qualitatifs de nos expériences, c'est-à-dire ce qu'on a appelé les *qualia*, ne le sont pas. Pour cela, nous n'avons pas besoin d'endosser l'hypothèse spéculative très controversée des « zombies », ces créatures qui sont nos répliques fonctionnelles mais qui seraient totalement dépourvues de conscience phénoménale¹⁸. Tout ce dont nous avons besoin, c'est de la possibilité du spectre inversé. Cette possibilité porte également à controverse, mais elle est manifestement moins problématique que celle des zombies.

Les émergentistes classiques de la première moitié du xx^e siècle ont soutenu que plusieurs aspects significatifs d'un tout complexe sont émergents, en ce sens que leur occurrence ne peut être ni prédite ni expliquée dans les termes des conditions sous-jacentes dont elles émergent. S'il doit y avoir au moins une chose à propos de laquelle ils ont vu juste, il semble bien que leur meilleure chance se trouve du côté des qualités phénoménales de l'expérience consciente. Que ces propriétés soient émergentes ou, sinon, réductibles à des traits biologiques ou physiques des organismes, cela dépendra, comme je l'ai soutenu, de la réponse à la question de savoir si elles peuvent ou non être expliquées comme étant des propriétés fonctionnelles définies par leurs rôles causaux. Cela est donc la question cruciale à laquelle nous devons répondre, afin d'être en mesure de répondre à la question plus générale de la réductibilité physique du mental. J'espère que, si nous n'avons pas réussi à répondre à cette dernière question de façon définitive, nous avons au moins réussi à bien la préciser¹⁹.

18. Pour une discussion plus détaillée de l'hypothèse des « zombies », voir Chalmers, D., *The Conscious Mind*, Oxford, Oxford University Press, 1996.

19. Ce texte a été traduit de l'anglais par Richard Vallée et Paul Bernier.

Réduction fonctionnelle et réduction logique

Max Kistler

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004938ar>
DOI : <https://doi.org/10.7202/004938ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Kistler, M. (2000). Réduction fonctionnelle et réduction logique. *Philosophiques*, 27 (1), 27–38. <https://doi.org/10.7202/004938ar>

Résumé de l'article

Kim attribue aux émergentistes un modèle de « réduction logique » dans lequel la prédiction ou l'explication d'une occurrence de la propriété réduite ne requiert, outre des informations sur le niveau réducteur, que des principes logiques et mathématiques. Sur la base de cette interprétation, je conteste deux thèses de Kim. La première concerne la légitimité du modèle émergentiste de réduction. J'essaie de montrer, à l'aide de l'exemple de l'addition des masses, que l'adoption de la réduction logique rendrait irréductibles certaines propriétés qui sont clairement réductibles. La deuxième est la thèse selon laquelle la réduction fonctionnelle correspond aux exigences émergentistes sur la réduction. Telle que Kim la caractérise, la réduction fonctionnelle comporte, outre une définition fonctionnelle de la propriété à réduire, l'indication des propriétés réalisatrices. Or cette information, qui correspond à la découverte d'une loi de correspondance (locale), est empirique et non seulement logique.

Réduction fonctionnelle et réduction logique

MAX KISTLER

kistler@ext.jussieu.fr

Département de Philosophie

Université Paris X — Nanterre

RÉSUMÉ. — Kim attribue aux émergentistes un modèle de « réduction logique » dans lequel la prédiction ou l'explication d'une occurrence de la propriété réduite ne requiert, outre des informations sur le niveau réducteur, que des principes logiques et mathématiques. Sur la base de cette interprétation, je conteste deux thèses de Kim. La première concerne la légitimité du modèle émergentiste de réduction. J'essaie de montrer, à l'aide de l'exemple de l'addition des masses, que l'adoption de la réduction logique rendrait irréductibles certaines propriétés qui sont clairement réductibles. La deuxième est la thèse selon laquelle la réduction fonctionnelle correspond aux exigences émergentistes sur la réduction. Telle que Kim la caractérise, la réduction fonctionnelle comporte, outre une définition fonctionnelle de la propriété à réduire, l'indication des propriétés réalisatrices. Or cette information, qui correspond à la découverte d'une loi de correspondance (locale), est empirique et non seulement logique.

ABSTRACT. — I read Kim as attributing to the emergentists a model of "logical reduction" in which the prediction or explanation of an occurrence of the reduced property requires, beyond information on the reducing level, only logical and mathematical principles. On the basis of this interpretation, I argue against two theses of Kim's. The first concerns the legitimacy of the emergentist model of reduction. I try to show by the example of the addition of masses that according to logical reduction certain properties which are clearly reducible come out as irreducible. The second is the thesis that functional reduction as construed by Kim satisfies the emergentist constraints on reduction. Functional reduction implies, over and above a functional definition of the reduced property, the indication of its realizers. But this information corresponding to the discovery of a (local) bridge law is empirical and not only logical.

Dans son texte récent « Émergence, modèles de réduction et le mental¹ », Kim ne nous offre pas moins qu'une remise en cause radicale de l'opinion reçue à la fois sur le modèle classique de la réduction proposé par Nagel et sur les conséquences de la conception fonctionnaliste des états mentaux concernant leur capacité ou non à être réduits à des états cérébraux². La conclusion du fameux argument de la réalisabilité multiple était qu'il est impossible de réduire les états mentaux à des états cérébraux. D'autre part, l'existence de cas de réduction réussie en l'absence de lois de correspondance (*bridge laws*) semblait montrer que le modèle nagelien était trop restrictif et qu'il fallait chercher des critères de réduction plus faibles que la découverte d'une loi de correspondance.

1. Voir l'article de Kim dans le présent numéro.

2. Pour le modèle classique de la réduction, voir Nagel, 1961.

Kim reconsidère ces arguments et parvient, dans les deux cas, à une conclusion inhabituelle. D'une part, il reconstruit un modèle de réduction qu'il attribue aux émergentistes du début du siècle et que j'appellerai le modèle de la « réduction logique ». Celui-ci impose des exigences plus fortes que celles du modèle nagelien, mais Kim fait valoir que le modèle est plausible et correspond à des intuitions largement répandues, entre autres chez les scientifiques. Ce n'est pas que Nagel soit trop exigeant, mais au contraire il ne le serait pas assez. D'autre part, contre la conviction répandue selon laquelle la conception fonctionnaliste des états mentaux entraîne leur irréductibilité, étant donné leur réalisabilité multiple, Kim élabore un concept de « réduction fonctionnelle » qui permet non seulement de montrer que le fonctionnalisme est compatible avec la réductibilité des états mentaux, mais qu'il s'agit d'une forme de réduction plus forte que celle du modèle nagelien : la réduction fonctionnelle serait une réduction logique au sens exigé par les émergentistes.

Il est toujours salutaire d'ébranler des vieilles certitudes. En ce sens, il faut saluer la démonstration de la modestie métaphysique de la réduction nagelienne qui la rend compatible avec des « théories dualistes classiques concernant la relation corps/esprit — comme par exemple la théorie spinoziste du double aspect, la doctrine de l'harmonie préétablie de Leibniz, le parallélisme ou le monisme neutre³ ». Il faut aussi reconnaître la plausibilité intuitive d'une doctrine de la réduction plus exigeante, comme celle de la réduction logique des émergentistes.

Dans ce qui suit, je conteste deux thèses cruciales avancées par Kim. La première concerne la légitimité du modèle émergentiste de réduction. J'essaie de montrer, en me servant de l'exemple de l'addition des masses, que l'adoption de la réduction logique rend émergentes, et donc irréductibles, certaines propriétés clairement réductibles. La deuxième est la thèse selon laquelle la réduction fonctionnelle satisferait aux exigences émergentistes sur la réduction.

1. Le modèle émergentiste de la « réduction logique »

Selon la reconstruction par Kim du concept de réduction employé par les émergentistes, une réduction nagelienne achevée ne constitue pas encore une réduction véritable, dans la mesure où ses prémisses ne font pas exclusivement référence aux propriétés et lois du niveau réducteur. La question-clé qui permet de distinguer les vraies réductions est la suivante : « L'occurrence d'une propriété de niveau supérieur peut-elle être prédite *uniquement* à partir d'informations concernant le niveau de base⁴? » Selon la réduction nagelienne, la réponse est non, car les lois de correspondance qui sont une composante essentielle de ses prémisses font référence à un lien entre propriétés

3. Voir Kim dans le présent numéro, p. 17

4. *Ibid.*

réductrices et propriétés réduites. Dans la mesure où les lois de correspondance font référence aux propriétés réduites — de haut niveau — la prédiction que fournit une réduction nagelienne, en accord avec le modèle déductif-nomologique (ou D-N), n'est pas basée sur la *seule* information concernant le niveau de base. Kim montre que cette exigence est l'expression d'une intuition largement répandue sur ce qu'il faut demander à une véritable réduction. Entre autres, son inclusion dans une définition de la réduction permettrait de justifier l'intuition selon laquelle les dualismes spinoziste, leibnizien ou cartésien ne devraient pas compter parmi les théories réductionnistes. Or, comme ces conceptions impliquent l'existence de lois de correspondance, elles comptent comme réductionnistes selon le modèle nagelien.

Mais une réduction sans lois de correspondance est-elle seulement envisageable? Si l'exigence émergentiste s'avérait trop forte pour être satisfaite, l'imposer serait un moyen de stipuler l'antiréductionnisme. Mais ce ne serait guère convaincant dans la mesure où cette conception déboucherait sur le résultat que toutes les propriétés macroscopiques sont irréductibles. Or une telle conclusion est inacceptable, étant donné qu'il existe au moins quelques cas non controversés de réduction réussie. L'intention des émergentistes eux-mêmes n'était pas de faire de *toutes* les propriétés macroscopiques des propriétés émergentes, mais au contraire d'opposer une classe particulière de propriétés émergentes à une autre classe, probablement plus vaste, de propriétés « résultantes ». Pour ces dernières, ils réclament la possibilité que leur occurrence, contrairement à celle des propriétés émergentes, puisse être prédite sur la *seule* base d'informations concernant le niveau de base.

En fait, une telle réduction, si elle ne présuppose pas d'informations empiriques sur le niveau à réduire, devrait néanmoins présupposer les principes de la logique et des mathématiques. Pour cette raison, je propose d'appeler « réduction logique » le modèle de réduction attribué par Kim aux émergentistes. Une propriété résultante est réductible en ce sens logique si sa prédiction est possible sur le modèle D-N, mais où les seules lois invoquées sont des principes logiques et mathématiques. La prédiction de la réduction logique est une sorte bien particulière de prédiction D-N, à savoir une prédiction « D-L » (pour « déductive-logique »), où les lois de correspondance à contenu empirique qui sont les prémisses *nomologiques* dans une prédiction D-N sont absentes, pour être remplacées par des principes purement logiques et mathématiques. En ce sens, une réduction logique justifierait l'affirmation que la propriété réduite « n'est rien d'autre que » sa base de réduction, où l'on sous-entendrait par « rien » : « rien d'empirique », mais seulement la logique et les mathématiques. Une réduction logique serait en même temps une réduction ontologique dans la mesure où une théorie ne contracterait aucun engagement ontologique en faveur de propriétés ainsi réduites, pour la raison qu'elles peuvent être logiquement déduites des seules propriétés du niveau réducteur. Ainsi, le modèle fait correctement apparaître le caractère non réductionniste des doctrines dualistes mentionnées plus

haut : ces conceptions sont incompatibles avec la réduction logique. Elles impliquent l'existence de liens nomologiques entre le mental et le physique, qui ne sont pas seulement logiques ou mathématiques.

Afin de tester la plausibilité de ce modèle, considérons la masse d'un objet complexe que Kim cite comme un cas paradigmatique de propriété « résultante » : l'information sur la masse d'un objet composé des parties o_1 et o_2 semble résulter de la seule information sur les masses m_1 et m_2 des composantes o_1 et o_2 . En ce sens, le calcul de la masse de l'objet complexe serait une prédiction D-L : rien en dehors des règles de l'addition ne semble requis pour déduire, à partir des masses m_1 et m_2 des composantes, que la masse du composé est $m_1 + m_2$. Comme Kim le fait remarquer dans une note, on ne considère plus aujourd'hui l'additivité de la masse comme un principe *a priori*, mais plutôt comme « empirique et contingent⁵ ». Le principe est en effet contingent car il est faux. La masse n'est additive qu'approximativement et non strictement, car dès que deux objets forment un tout par l'établissement d'un lien physico-chimique, la masse de l'ensemble est plus petite que la somme des masses, par la quantité correspondant à l'énergie de liaison.

Or si le rapport entre les masses des parties et la masse du tout est empirique (connue seulement *a posteriori*), Kim a tort d'affirmer que « nous pouvons expliquer pourquoi la table a cette masse à partir d'informations à propos de la masse de ses parties⁶ », où il est sous-entendu que cette explication ne fait appel à aucune loi ni à d'autres faits empiriques en dehors des masses des deux parties. Cela ressort de l'affirmation que « la masse de cette table peut être dérivée, logiquement ou mathématiquement, de la masse de chacune de ses parties, disons le dessus de la table et sa structure⁷ ». Or dès qu'on admet l'idée que l'additivité des masses n'est pas connue *a priori*, mais obéit à une loi empirique qui dépend du type de lien entre les parties lorsqu'elles sont assemblées, la dérivation de la masse du tout devient impossible sans faire appel à cette loi⁸.

5. *Ibid.*, p. 18 note 11.

6. *Ibid.*, p. 18.

7. *Ibid.*

8. La masse M d'un objet complexe est certainement une propriété basée microscopiquement (micro-based) au sens où le fait d'avoir M « est la propriété d'avoir des parties propres $a1, a2, ..., an$, tels que $P1(a1), P2(a2), ..., Pn(an)$, et $R(a1, ..., an)$ » (Kim, 1997a, p. 292), où $P1$ est la propriété d'avoir la masse de la partie $a1$, etc. Rien (à l'éventuelle exception des particules élémentaires) ne peut avoir une masse M sans avoir des parties qui ont une masse. Mais on ne peut calculer M à partir de la seule masse de ses parties. Il faut tenir compte également de l'effet nomique de la relation R sur le résultat de cette « addition physique » de masses qui, en général, n'est pas simplement l'addition mathématique. Comme le dit Kim ailleurs, « la détermination n'implique, par elle-même, ni la prédictabilité ni l'explicabilité » (Kim, 1997a, p. 296). On peut interpréter cela en faisant remarquer que la détermination peut être partielle : les masses des parties déterminent partiellement, mais non totalement, la masse du tout. Mais la masse du tout ne serait prédictible sur la seule base de la masse des parties que si la détermination était totale.

Faut-il généraliser ce constat pour conclure qu'il en est ainsi de toutes les propriétés macroscopiques : qu'elles sont toutes « émergentes » au sens où elles sont « logiquement irréductibles », c'est-à-dire qu'il est impossible de prédire sur la base de la seule information sur les propriétés des parties, plus la logique et les mathématiques, la moindre propriété du tout? Pour rester avec l'exemple de la masse, il semble bien possible de prédire, sur la base de la seule logique, que le tout a *une* masse, même si on ne sait pas laquelle. Mais en réalité, même cette prédiction bien modeste est fondée sur un principe empirique. L'interaction de certaines particules massives mène à la transformation d'une partie de leur masse en rayonnement. Nous ne pouvons savoir qu'*a posteriori* dans quelles circonstances cela arrive et dans lesquelles cela n'arrive pas, et nous ne pouvons savoir qu'*a posteriori* que si une transformation de masse en énergie a lieu elle ne concerne qu'une partie de la masse, et donc qu'un tout résultant de l'interaction de deux parties d'un certain type a encore une masse quelconque, plutôt que d'avoir entièrement été transformé en rayonnement.

Il n'est heureusement pas nécessaire de poursuivre plus loin la question de savoir s'il existe des macropropriétés résultantes pour montrer que la conception émergentiste de la réduction est inutilisable. Car il est difficile de trouver une propriété dont la réductibilité est moins douteuse que celle de la masse d'un objet complexe par rapport aux masses de ses parties. Donc, si la réduction logique fait de cette propriété une propriété irréductible — selon le critère énoncé plus haut de la dérivabilité sur la seule base de connaissances *a priori* jointes aux informations sur les parties — elle ne peut servir de modèle de la réduction telle que nous la connaissons intuitivement.

L'argument de la réalisabilité multiple montre que Nagel avait tort de penser qu'il est nécessaire pour la réduction d'une propriété et des lois dans lesquelles elle participe qu'il existe des lois de correspondance⁹. Mais le modèle émergentiste de la réduction commet l'erreur opposée : ce n'est pas parce qu'en général il n'existe pas de loi de correspondance empirique permettant de déduire la possession d'une propriété macroscopique de la possession microscopique qu'il faut conclure que les réductions sont possibles sans aucun recours à des faits empiriques quelconques, c'est-à-dire sur la seule base de la logique et des mathématiques.

2. La réduction fonctionnelle

Avec la « réduction fonctionnelle¹⁰ », Kim estime avoir trouvé un modèle de réduction qui doit être préféré à celui de Nagel et qui est une « réduction

9. Dans Kistler, 1999, le cas de la température est analysé pour montrer que la réalisabilité multiple est compatible avec la réductibilité d'une propriété.

10. Kim, 1997b, introduit l'idée de la « fonctionnalisation » d'une propriété comme étape décisive vers sa réduction.

logique » au sens où il est conforme à l'exigence émergentiste sur la réduction : la dérivabilité logique sur la seule base des propriétés et lois du niveau de la base de réduction. Mais j'essaierai de montrer que la réduction fonctionnelle n'est pas une réduction logique.

La réduction fonctionnelle comporte trois étapes : premièrement, la propriété « macroscopique » *M* reçoit une définition fonctionnelle en termes du rôle causal qu'elle joue. Cette définition prend la forme d'une quantification existentielle sur les propriétés du niveau de base qui occupent ce rôle grâce à leurs liens causaux/nomiques à d'autres propriétés du niveau de base¹¹. Lors de la deuxième étape, cruciale pour notre argument, on découvre la propriété *P* qui réalise la propriété fonctionnelle *M* dans les individus de l'espèce *S*, où *P* est une propriété microscopique relativement à *M*.

Comme nous le supposons dans ce cas, une réduction fonctionnelle a été accomplie, et nous supposons donc comme étant acquise la connaissance du réalisateur de la propriété *M* dans des systèmes comme *S*, à savoir *P* qui est une propriété ou un mécanisme au niveau de base¹².

La troisième étape consiste à élaborer une théorie qui permet de comprendre comment la propriété *P* parvient à jouer le rôle causal indiqué dans la première étape¹³.

Il est vrai qu'il est possible, une fois les trois étapes accomplies, de prédire des occurrences de la propriété *M*, sur la base des seules informations portant sur le niveau de base. Si *P* est une propriété neurologique, il sera possible de prédire une occurrence de *M* à partir de l'observation de processus neurophysiologiques qui donnent régulièrement lieu à l'occurrence de *P*. Mais l'enjeu est de savoir si cette prédiction est possible sur cette seule base. C'est ce qu'affirme Kim lorsqu'il dit qu'après une réduction fonctionnelle « les occurrences de *M* peuvent effectivement être prédites à partir d'informations concernant uniquement le niveau de base¹⁴ » — et c'est uniquement la vérité de cette affirmation qui pourrait justifier la thèse selon laquelle la réduction fonctionnelle est une réduction logique. Mais cette affirmation est fausse. Examinons la deuxième étape de la réduction fonctionnelle, qui est cruciale parce que c'est là qu'est introduit le lien entre les niveaux de *M* et de *P*. Kim reconnaît que le lien entre propriété réalisatrice et propriété réalisée n'est pas d'ordre logique : « Le fait qu'une propriété neuronale donnée

11. Voir Kim, 1997b.

12. Voir Kim dans le présent numéro, p. 21

13. Cette façon de présenter les choses montre que la réduction fonctionnelle est un processus itératif : on peut concevoir la propriété *P* à son tour comme caractérisée de manière fonctionnelle dans le réseau nomologique auquel elle prend part. Cela permet de la considérer à son tour comme une propriété macroscopique relativement à un niveau encore inférieur en taille et complexité auquel elle est susceptible d'être réduite, dès qu'elle trouve sa ou ses propriété(s) réalisatrice(s).

14. Voir Kim dans le présent numéro, p. 21.

réalise la douleur, par exemple, est quelque chose d'empirique et de contingent¹⁵ ». Faisons abstraction, dans le présent contexte, de la thèse de la contingence¹⁶. Dire que le fait que M soit réalisé par P dans l'espèce S est empirique revient à dire que le lien de réalisation est connu *a posteriori*.

Pour que le lien de réalisation entre P et M dans S puisse jouer un rôle dans l'explication et la prédiction, il faut que ce lien soit nomologique et non seulement factuel. Il faut que les individus de l'espèce S possèdent la propriété P à chaque fois qu'ils possèdent la propriété M , non seulement par accident mais en vertu d'une loi, faute de quoi on ne pourrait ni expliquer ni prédire qu'un individu donné d'espèce S possède M , à partir de la prédiction neurophysiologique qu'il possède P . La prédiction ou l'explication que l'individu a M nécessite la prémisse nomologique selon laquelle P réalise M dans l'espèce S ; autrement dit, elle nécessite une loi de correspondance locale. Ailleurs, Kim a explicitement présenté la réduction accomplie par la découverte de la propriété réalisatrice d'une propriété qui avait d'abord été caractérisée fonctionnellement, comme une « réduction locale » qui passe par « un ensemble de lois de correspondance restreintes à S , de la forme $S \rightarrow (M_i \ P_i)$, pour chaque espèce mentale M_i ¹⁷. Lors de la réduction locale, qui est donc plus faible que l'exigence nagelienne de réduction globale, une propriété de niveau plus élevé n'est pas *simplement* Nagel-réduite à une propriété de niveau plus bas, mais elle est Nagel-réduite *relativement* à chacun des domaines d'un nombre de domaines S disjoints.

Dès qu'on reconnaît l'intervention essentielle des lois de correspondance locales dans la réduction fonctionnelle, on ne peut plus la considérer comme une réduction logique¹⁸. Si les réductions fonctionnelles nécessitent des lois de correspondance locales, il n'est plus vrai que « celles-ci sont *réductrices* en ce sens que les ressources explicatives et prédictives qui sont mises à profit proviennent exclusivement du niveau réducteur de base¹⁹ ». Kim justifie l'affirmation d'« exclusivité » en disant que :

15. *Ibid.*, p. 24.

16. Pour montrer que la réduction fonctionnelle n'est pas une réduction logique, il suffit de reconnaître que le lien n'est pas logique et donc *a priori* mais empirique et donc *a posteriori*. En outre, il ne s'agit pas d'un simple fait contingent mais plutôt d'un lien nomologique. Je serais en désaccord avec Kim sur l'interprétation du statut modal des lois — contingentes ou nécessaires. Mais il n'est pas nécessaire de trancher ce débat ici. Le point important et non controversé est que les lois ne sont connues qu'*a posteriori*.

17. Voir Kim, 1992, p. 21. Dans le même sens, Kim dit qu'« il suit de la définition de la réalisation que P est suffisant pour M — en fait, étant donné la constance nomologique [...] de la relation de réalisation, il suit que P est nomologiquement suffisant pour M » (Kim, 1997b, p. 197).

18. Kim note explicitement que la survenance impliquée par la relation de réalisation qui est elle-même nomologique et non logique, a la force de la « nécessité nomologique (plutôt que logique/métaphysique) » (Kim, 1997b, p. 197).

19. Voir Kim dans le présent numéro, p. 21.

C'est une définition qui nous fournit les relations réductrices entre le niveau de base et la propriété cible, en l'occurrence une caractérisation fonctionnelle de M en termes de relations causales impliquant des éléments du niveau de base²⁰.

Mais la définition fonctionnelle de M ne fournit qu'une quantification existentielle sur des propriétés du niveau de base qui satisfont une certaine spécification causale/nomique²¹. Une telle quantification existentielle ne fournit pas de lien déductif du niveau de base au niveau de M dans le sens nécessaire pour la prédiction de l'occurrence de M , à partir des seules informations portant sur le niveau de base. L'argument suivant n'est pas déductivement valide :

(1) Fa , où a est un organisme capable d'exemplifier des propriétés mentales en exemplifiant une propriété réalisatrice du niveau de base (neurophysiologique), et F est une propriété du niveau de base.

(2) $(\forall x)(Mx \supset (\exists P)(Px \ \& \ D(P)))$, où D est une condition causale/nomique sur les propriétés P du niveau de base²².

Donc, Ma .

Il est vrai que (2), la définition fonctionnelle de M , fournit un lien entre les deux niveaux de propriétés, et même un « lien déductif », mais en l'absence d'information sur le fait que la propriété F du niveau de base figure parmi les réalisatrices de M , il est impossible de déduire l'occurrence de M à partir d'information sur l'occurrence de F . Il est également vrai qu'il n'est pas nécessaire, pour prédire l'exemplification de M par a , de connaître la propriété particulière qui réalise M dans les organismes du type de a . Il suffit de compléter les prémisses (1) et (2) ou bien par l'ensemble des propriétés réalisatrices de M (parmi lesquelles figure F), ou bien par un sous-ensemble de ces réalisatrices comprenant F . Mais en tout état de cause, la seule quantification existentielle figurant dans la définition fonctionnelle (2) n'y suffit pas.

Notre thèse selon laquelle la deuxième étape d'une réduction fonctionnelle mobilise une relation de réalisation *nomologique* (une loi de correspondance locale²³) est plus forte qu'il n'est nécessaire pour le présent argument qui vise à montrer que la réduction fonctionnelle n'est pas une réduction *logique* : pour cela, il suffit de montrer que le lien de réalisation entre M et P dans S est un fait empirique. Or, comme nous l'avons noté plus haut, cela est explicitement reconnu par Kim²⁴.

20. *Ibid.*, p. 21.

21. Voir Kim, 1997b, p. 195.

22. Cette équivalence correspond à la conception fonctionnaliste des propriétés mentales comme propriétés de second ordre, et plus précisément des « propriétés fonctionnelles » qui sont les « propriétés de second ordre sur B [soit une classe de propriétés non mentales du niveau de base, qui inclut des propriétés physico-chimiques, biologiques et comportementales] dont la spécification D implique la relation causale/nomique » (Kim, 1997b, p. 194-195).

23. Cette thèse peut s'appuyer sur certaines suggestions de Kim ; voir Kim, 1992 et 1997b.

24. Voir la note 15.

Kim nie l'importance de la deuxième étape de la réduction fonctionnelle qui, selon notre analyse, est cruciale et qui consiste à trouver la propriété réalisatrice P d'une propriété M , pour un système de type S . Pour Kim « la question cruciale » est de savoir si la propriété est « fonctionnalisable », ce qui exige qu'elle puisse recevoir une définition implicite fonctionnelle et que son rôle fonctionnel soit rempli par une propriété réalisatrice quelconque. Et, comme il le note, « si une propriété mentale a été fonctionnalisée, le reste de la tâche revient à la science et, d'un point de vue métaphysique, c'est comme si elle avait été réduite²⁵ ». Or, si notre analyse est correcte, la fonctionnalisation seule ne suffit pas pour l'explication ou la prédiction d'une occurrence de M : il faut en outre connaître la loi de correspondance locale $S \rightarrow (M \ P)$ ²⁶. Dans ses propres termes, on peut exprimer cette conclusion ainsi : « La possibilité de la fonctionnalisation [...] est une condition nécessaire de la microréduction », mais elle n'est pas suffisante, car « la question de savoir si la réduction sera actuellement accomplie est une question empirique [...] de la science²⁷ ». La réduction fonctionnelle, c'est-à-dire l'indication d'une définition fonctionnelle, ou fonctionnalisation, ne mène qu'à la réductibilité²⁸, mais non à la réduction, et en particulier pas à la réduction logique.

Par conséquent, l'argument par lequel Kim montre que la réduction nagelienne n'est pas une réduction logique s'applique *mutatis mutandis* à la réduction fonctionnelle qui est elle aussi nagelienne, quoique seulement de façon locale. À la place de la prédiction nagelienne suivante :

« Le système S exemplifiera T à t

$M \ T$

Donc, S exemplifiera M à t ²⁹ »,

où « $M \ T$ » est une loi de correspondance globale, la réduction fonctionnelle fournit la prédiction suivante :

Le système S exemplifiera T à t

$S \rightarrow (M \ T)$

Donc, S exemplifiera M à t »,

25. Voir Kim le présent numéro, p. 24.

26. Ce que la fonctionnalisation établit à elle seule, c'est que la propriété fonctionnalisée est causalement efficace, en héritant des pouvoirs causaux de ses propriétés réalisatrices ; voir Kim, 1997a, p. 295. Cela s'explique du fait que fonctionnaliser M signifie définir M en termes des relations causales dans lesquelles entrent les objets qui possèdent M , en vertu du pouvoir causal de l'une de leurs propriétés réalisatrices.

27. Voir Kim, 1997b, p. 202-203.

28. « La conception fonctionnaliste des propriétés mentales est [...] nécessaire et suffisante pour la réductibilité » (Kim, 1997b, p. 203).

29. Voir Kim dans le présent numéro, p. 22.

où « $S \rightarrow (M \ T)$ » est une loi de correspondance locale qui dit que dans les systèmes S , M est réalisée par T .

On peut donc retourner la critique kimienne de la réduction nagélienne contre sa propre réduction fonctionnelle. La citation suivante de Kim s'applique aussi bien aux prédictions du second type qu'à celles du premier type : « De telles prédictions d'occurrences de M ne sont donc pas fondées *uniquement* sur la base d'informations concernant le phénomène et les lois au niveau de base et, par conséquent, elles ne satisfont pas le *desideratum* sur la réduction³⁰ », c'est-à-dire la réduction logique.

On peut également retourner contre Kim la critique qu'il adresse à la théorie selon laquelle les propriétés mentales sont identiques à des propriétés cérébrales à la manière des identités nécessaires mais *a posteriori* de Kripke, sur le modèle de l'identité eau = H₂O, à savoir :

Comme ces identités [...] ne peuvent être connues qu'empiriquement, et comme elles sont requises pour toute prédiction d'occurrences mentales reposant sur de l'information relative aux neurones, par exemple pour la prédiction d'occurrences de la douleur, il ne sera pas possible de prédire de telles occurrences *uniquement* sur la base d'information neurobiologique ou behaviorale à propos des organismes³¹.

Étant donné que le lien factuel (et nomologique) entre la propriété fonctionnelle M et sa réalisatrice dans un individu de l'espèce S ne peut lui aussi être connu qu'empiriquement, et que la connaissance de ce lien est indispensable à la prédiction d'une occurrence de M à partir d'informations portant sur le niveau de base, cette prédiction n'est pas non plus conforme aux exigences de la réduction logique.

3. Conclusion

La doctrine émergentiste cherche à rendre justice à l'intuition selon laquelle la satisfaction des contraintes nagéliennes sur la réduction, loin de suffire à une réduction achevée, ne permet au mieux que de poser clairement la question suivante, à laquelle la réduction répondrait : pourquoi la propriété M est-elle nomologiquement liée, dans les individus de l'espèce S , à la propriété réalisatrice P ? Considérer M comme une propriété émergente, aussi longtemps qu'aucune réponse n'est trouvée et que ce lien reste un fait brut inexpliqué, c'est donner expression à l'intuition que l'existence de M est, dans ces conditions, inintelligible du point de vue physicaliste. Kim propose la réduction fonctionnelle comme un modèle de réduction qui ne permet plus seulement de poser la question mais d'y répondre. Donner une définition fonctionnelle de la propriété M , c'est la caractériser en termes de relations

30. *Ibid.*

31. *Ibid.*, p. 24.

causales/nomiques que les objets qui possèdent M entretiennent grâce à leur possession de l'une ou l'autre des propriétés réalisatrices de M . Comme la propriété réalisatrice est mentionnée dans la *définition* même de M , le fossé entre le niveau réalisateur et le niveau réalisé semble être comblé par un lien conceptuel et donc *a priori* intelligible. Par conséquent, Kim affirme avoir trouvé, avec la réduction fonctionnelle, une forme de réduction qui répond à l'exigence émergentiste de rendre le lien entre la propriété M et sa propriété réalisatrice intelligible, plutôt que de le laisser à l'état de fait brut mystérieux comme le fait la réduction nagelienne.

Mais, j'ai cherché à montrer que la réduction fonctionnelle permet d'expliquer ou de prédire une occurrence de M uniquement grâce à l'introduction d'un nouveau fait brut qui, du point de vue émergentiste, est tout aussi mystérieux qu'une loi de correspondance nagelienne qui, tout comme cette dernière, est un fait nomologique et qui dépasse la seule définition fonctionnelle ; il s'agit du fait empirique que M possède les propriétés réalisatrices $P_1, P_2, \dots P_n$. La question n'est plus : pourquoi M est-elle liée à P par une loi de correspondance ? mais plutôt : pourquoi M est-elle réalisée par P_1 ou P_2 ou $\dots P_n$? Mais il n'y a de réponse purement logique ou mathématique ni à l'une ni à l'autre de ces questions, et en ce sens, selon le critère émergentiste, M resterait toujours une propriété émergente même après une réduction fonctionnelle.

Il n'y a pas lieu de contester l'intuition qui supporte le critère émergentiste de la réduction logique. Il semble parfaitement légitime de trouver qu'il existe un sens dans lequel même la découverte du réalisateur P de la douleur dans les humains ne nous permet toujours pas de comprendre d'où vient la douleur, sur la base de la connaissance de P et des processus qui mènent aux occurrences de P . Après tout, comme j'ai essayé de le montrer, la seule connaissance du niveau de base ne suffit pas pour déduire logiquement une occurrence de douleur. Mais la considération du cas de la masse d'objets complexes plaide contre l'interprétation émergentiste de la réduction adoptée par Kim, c'est-à-dire contre l'idée d'identifier les propriétés émergentes et irréductibles et les propriétés résultantes et réductibles. La masse d'un objet complexe ne se laisse pas déduire logiquement ou mathématiquement à partir de la seule information sur ses parties et leurs masses, sans information sur la loi de correspondance qui spécifie la masse du tout étant donné les masses des parties et la nature de leur liaison. Le concept émergentiste de « réduction logique » fait de la masse du tout une propriété irréductible (car émergente) à la masse des parties. Mais un concept de réduction qui mène à considérer comme irréductible une propriété aussi clairement réductible que la masse d'un objet complexe est trop étroit.

Bibliographie

- Kim, J., 1992. « Multiple Realization and the Metaphysics of Reduction », *Philosophy and Phenomenological Research*, 52, p. 1-26 ; réédité dans Kim, J., *Supervenience and Mind*, Cambridge (Mass.), Cambridge University Press, 1993, chap. 16.
- Kim, J., 1997a. « Does the Problem of Mental Causation Generalize? », *Proceedings of the Aristotelian Society*, 97, p. 281-297.
- Kim, J., 1997b. « The Mind-Body Problem : Taking Stock after Forty Years », dans Tomberlin, J, dir., *Philosophical Perspectives 11, Mind, Causation, and World*, Boston, Basil Blackwell, p. 185-207.
- Kim, J., 2000. « Émergence, modèles de réduction et le mental », *Philosophiques*, 27, n° 1, p. 9-24.
- Kistler, M., 1999. « Multiple Realization, Reduction, and Mental Properties », *International Studies in the Philosophy of Science*, 13, n° 2, p. 135-149.
- Nagel, E., 1961. *The Structure of Science*, London, Routledge and Kegan Paul.

L'empire contre-attaque : le retour de la réduction psychophysique

Pierre Poirier

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004969ar>
DOI : <https://doi.org/10.7202/004969ar>

[Aller au sommaire du numéro](#)

Éditeur(s)
Société de philosophie du Québec

ISSN
0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Poirier, P. (2000). L'empire contre-attaque : le retour de la réduction psychophysique. *Philosophiques*, 27 (1), 39–62. <https://doi.org/10.7202/004969ar>

Résumé de l'article

En refusant à la psychologie la latitude accordée aux autres sciences, l'argument concluant à l'irréductibilité des propriétés psychologiques à partir de leur réalisation multiple manifeste une attitude antinaturaliste à l'égard de cette science. En science, il est possible de relativiser les réductions à des domaines bien définis, c'est-à-dire des domaines qui découpent la nature d'une manière non *ad hoc*, et de corriger en conséquence l'appareil conceptuel des théories. Et en science, il est possible de construire des niveaux abstraits et idéalisés permettant la description du comportement global des systèmes, niveaux qui font abstraction de complexités inutiles. Si l'on accorde les mêmes privilèges à la psychologie, la réalisation multiple des propriétés psychologiques ne permet pas d'inférer leur irréductibilité.

L'empire contre-attaque : le retour de la réduction psychophysique

PIERRE POIRIER

poirier_pierre@ucdavis.edu

Philosophy Department

University of California at Davis

RÉSUMÉ. — En refusant à la psychologie la latitude accordée aux autres sciences, l'argument concluant à l'irréductibilité des propriétés psychologiques à partir de leur réalisation multiple manifeste une attitude antinaturaliste à l'égard de cette science. En science, il est possible de relativiser les réductions à des domaines bien définis, c'est-à-dire des domaines qui découpent la nature d'une manière non *ad hoc*, et de corriger en conséquence l'appareil conceptuel des théories. Et en science, il est possible de construire des niveaux abstraits et idéalisés permettant la description du comportement global des systèmes, niveaux qui font abstraction de complexités inutiles. Si l'on accorde les mêmes privilèges à la psychologie, la réalisation multiple des propriétés psychologiques ne permet pas d'inférer leur irréductibilité.

ABSTRACT. — In science, reductions can be relative to specific, well-defined domains, domains that carve nature in a non *ad hoc* way. Relativization to domains is a time-honored way to adjust the conceptual apparatus of theories. And in science, schemes can be defined to describe the global behavior of systems in a way that abstracts from unnecessary (and annoying) detail. I show that the classical argument from multiple realizability to non-reducibility vanishes once the same leeway is granted to psychology. By denying psychology the use of these standard theory-building strategies in science, the argument exhibits an antinaturalist attitude towards psychology, an attitude that may be welcome in some antiphysicalist quarters but that should be shunned in cognitive science.

La question de la réduction du psychique au physique est au cœur du problème métaphysique de l'esprit. La théorie classique de la réduction théorique, dont on retrouve la version canonique dans *The Structure of Science* de Nagel¹, devait toutefois remplir une fonction purement épistémique, soit celle d'explicitier les conditions sous lesquelles on peut dire d'une réduction donnée qu'elle est *justifiée*. Ce fait ne devrait surprendre personne : la théorie classique fut développée par les partisans d'une philosophie franchement hostile à la métaphysique. L'analyse néopositiviste des propriétés formelles et sémantiques générales du discours montrait que la métaphysique ne pouvait en aucun cas se prononcer sur les composantes du réel et qu'elle devait être rejetée comme discours insensé lorsqu'elle prétendait le faire. Qui pensait le contraire confondait tout simplement les modes formels et matériels du discours. Cependant, avec la publication coup sur coup des articles de Feigl, « The "Mental" and the "Physical" », et de Smart, « Sensations and Brain

1. Nagel, E., *The Structure of Science*, New York, Harcourt, Brace & World, 1961.

Processes », puis des réponses qu'ils ont suscitées, pensons notamment aux articles de Putnam, « Psychological Predicates », et de Davidson, « Mental Events² », la philosophie se redonnait le droit de traiter de questions métaphysiques et donnait par le fait même naissance au problème de l'esprit (*mind-body problem*) tel que nous le connaissons aujourd'hui³. Du même coup, la théorie classique de la réduction acquérait une fonction métaphysique : déterminer les constituants ultimes du monde en *faisant connaître*, en anticipation du travail scientifique, quelles théories possèdent les propriétés formelles et sémantiques nécessaires pour que leur réduction à des théories dont nous acceptons déjà les engagements ontologiques soit possible.

Avec la chute du mouvement néopositiviste, l'intérêt des philosophes s'est ainsi déplacé de la justification de la réduction à la détermination de la réductibilité. C'est dans cet esprit que des philosophes favorables à l'idée d'une psychologie intentionnelle qui soit scientifique ont opté pour l'argument antiréductionniste fondé sur la thèse de la réalisation multiple⁴. Il est important de noter que, contrairement à l'usage originel de la théorie classique de la réduction, où le travail philosophique (de justification) est postérieur au travail scientifique à proprement parler, ces nouveaux usages « par anticipation » et métaphysiques de la théorie sont à la merci des développements scientifiques. Ils ne valent que dans la mesure où la science du jour permet une projection *fiable* vers la science de demain. Dans cet article, je montrerai que des développements récents en neurosciences mathématiques (*computational neuroscience*) nous permettent d'ores et déjà d'affirmer que l'argument antiréductionniste fondé sur la réalisation multiple des états psychiques n'est pas valide. Si mon argument est valide, il faudra conclure que la psychologie intentionnelle peut se réduire à la neurologie. Dire qu'elle *peut* se réduire ne signifie pas, évidemment, qu'elle se réduira *de fait*, ou qu'elle se réduira *en entier* ou encore qu'elle se réduira *telle quelle*. Cela signifie, cependant, qu'une position antiréductionniste *a priori* est intenable.

Quelques philosophes ont récemment proposé des nouveaux modèles de la réduction⁵ qui, pour des raisons propres à chacun qu'il ne convient pas de

2. Feigl, H., « The "Mental" and the "Physical" », *Minnesota Studies in the Philosophy of Science*, 2, 1958, p. 370-497 ; Smart, J. J. C., « Sensations and Brain Processes », *Philosophical Review*, 68, 1959, p. 141-156 ; Putnam, H., « Psychological Predicates », 1967, réédité sous le titre « The Nature of Mental States », dans Putnam, H., *Mind, Language, and Reality*, Cambridge, Cambridge University Press, 1975 ; Davidson, D., « Mental Events », 1970, réédité dans Davidson, D., *Essays on Action and Events*, Oxford, Oxford University Press, 1980.

3. Kim, J., *Mind in a Physical World*, Cambridge (Mass.), MIT Press, 1998.

4. Putnam, « Psychological Predicates » ; Fodor, J. A., « Special Sciences », *Synthese*, 28, 1974, p. 97-115.

5. Bechtel, W. et McCauley, R. N., « Heuristic Identity Theory (or Back to the Future) : The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience », communication présentée au 25^e congrès annuel de la *Society for Philosophy and Psychology*, Palo Alto, 1999 ; Bickle, J., *Psychoneural Reduction*, Cambridge (Mass.), MIT Press, 1998 ; Kim, *Mind in a Physical World*.

reprennent ici, brisent l'inférence allant de la réalisation multiple des états psychiques à leur irréductibilité. J'affaiblirai davantage la position antiréductionniste en montrant que, contrairement à ce que l'on prétend depuis trente ans, elle ne découle pas non plus de la conception *classique* de la réduction et qu'il n'est donc pas nécessaire de construire de nouveaux modèles de la réduction pour contrer l'argument de la réalisation multiple. Je ne veux pas laisser entendre par là que ces nouveaux modèles sont construits uniquement pour résoudre ce problème et qu'ils seraient par conséquent inutiles à la lumière des arguments du présent article. Ces modèles traitent aussi d'autres problèmes qui ne trouvent peut-être pas de solution dans la théorie classique de la réduction, pensons par exemple au problème de la causalité mentale.

1. Le fonctionnalisme et l'argument antiréductionniste fondé sur la réalisation multiple

Nous allons reprendre en détail l'argument de la réalisation multiple des états psychiques, car il sera important de préciser la prémisse que nous voulons rejeter. Il est parfois professé et souvent sous-entendu par certains philosophes fonctionnalistes que le fonctionnalisme entraîne l'irréductibilité du psychique, et donc que l'adoption d'une conception fonctionnaliste implique le rejet du réductionnisme⁶. Mais comme le notent Bickle et Kim, l'avantage du fonctionnalisme est qu'il peut s'accommoder d'une variété de positions ontologiques au sujet de l'esprit (y compris du dualisme des substances de Descartes)⁷. Le fonctionnalisme permet l'antiréductionnisme mais ne le recommande ni ne l'impose. Il permet la définition d'une conception cohérente du domaine psychique qui ne soit pas réductionniste, mais cela ne signifie pas que sa vérité dépende de la fausseté du réductionnisme ou qu'elle l'implique⁸. Puisque le fonctionnalisme peut s'accommoder aussi bien du réductionnisme que de l'antiréductionnisme, il faut donc le dissocier de l'une ou l'autre de ces positions. Dans la présente section, nous rappelons comment le fonctionnalisme permet la formulation d'une conception de l'esprit qui soit compatible avec la thèse de la réalisation multiple des états psychiques. Nous soulignons ensuite que bien qu'il soit raisonnable de croire que les états psychiques sont bel et bien réalisés de diverses façons, cela n'est pas de nature à empêcher la formulation des lois de correspondance nécessaires à la justification d'une réduction des théories psychologiques, contrairement à ce que l'on professe.

Un matérialiste peut maintenir que la psychologie offre une description abstraite des états et processus cérébraux servant d'intermédiaires entre la

6. Voir notamment Putnam, « Psychological Predicates », et Fodor, « Special Sciences ».

7. Bickle, *Psychoneural Reductions* ; Kim, *Mind in a Physical World*.

8. L'étape fondamentale de la théorie de la réduction de Kim, *ibid.*, demande par exemple la « fonctionnalisation » des propriétés à réduire.

sensation et le comportement. Mais comment cette psychologie individuerait-elle ces *états abstraits* du cerveau? Une réponse précise à cette question a dû attendre certaines avancées en sémantique des prédicats théoriques, dont nous nous contentons ici de présenter les grandes lignes : pour toute théorie scientifique T , (1) la valeur sémantique de T dépend de son degré de confirmation ; (2) la valeur de vérité de chaque énoncé de T dépend de celle de tous les autres ; (3) la signification des termes de T dépend de leur rôle dans T ; chaque terme est ainsi *implicitement défini* par la théorie. David Lewis a formalisé cette nouvelle conception de l'interprétation des termes théoriques⁹. Si T est une théorie scientifique confirmée, il est alors possible de l'exprimer au moyen d'un énoncé de la forme « $\exists x_1, \dots, \exists x_n T(x_1, \dots, x_n)$ » où $x_1, \dots, x_n$ sont des variables tenant la place des prédicats de la théorie (transformés en termes singuliers). Au moyen de ces « énoncés de Ramsey », nous pouvons alors dire que le prédicat P_1 de T , qui correspond à la variable x_1 de T « ramseyifiée », exprime la propriété d'être une chose du type de celles qui jouent le rôle représenté par x_1 dans T ramseyifiée. La théorie décrit ainsi la fonction ou le « rôle fonctionnel » joué par les objets constituant l'extension de ses prédicats. P_1 exprime la propriété d'être une chose d'un type ayant un certain rôle fonctionnel, *et tout objet capable de remplir ce rôle fera partie de l'extension du prédicat*. Un rôle fonctionnel est donc une propriété de types d'individus caractérisée par la place qu'ils occupent au sein d'une théorie. Par extension, les propriétés exprimées par de tels prédicats sont appelées « fonctionnelles ». Parce que n'importe quel objet capable de remplir un rôle fonctionnel décrit par une théorie aura par le fait même la propriété fonctionnelle, il s'ensuit que tout prédicat fonctionnel peut, en principe, être réalisé de diverses façons. Cette définition de la nature des prédicats fonctionnels ouvre ainsi la possibilité de leur réalisation multiple.

Cette conception de l'interprétation des termes théoriques s'applique naturellement à la psychologie. Tout comme le prédicat P_1 est implicitement défini par les lois constituant T , les prédicats psychologiques comme « croit (que p) » seraient implicitement définis par le réseau de lois ou de principes pratiques constituant l'appareil nomologique de la psychologie du sens commun. Les lois ou principes pratiques de cette psychologie détermineraient un rôle fonctionnel pour chaque prédicat exprimant une de ses catégories. Un prédicat « mentaliste » exprimerait ainsi une propriété fonctionnelle : la propriété d'être une chose du type de choses jouant tel rôle fonctionnel déterminé par la psychologie du sens commun. Et, tout comme précédemment, il s'ensuit que tout état (neurologique, électronique, etc.) capable d'occuper le rôle approprié est un état psychique. La conception fonctionnaliste permet ainsi d'assigner une signification aux prédicats « mentalistes » malgré l'existence

9. Voir Lewis, D., « How to Define Theoretical Terms », *The Journal of Philosophy*, 67, 1970, p. 427-446 ; Lewis, D., « Psychophysical and Theoretical Identifications », *Australasian Journal of Philosophy*, 50, 1972, p. 249-258.

de multiples référents ne possédant rien d'autre en commun que leur capacité de remplir un certain rôle fonctionnel spécifié par la théorie.

Permettre formellement une chose, ce n'est toutefois pas militer en faveur de son existence. Pour tirer un argument antiréductionniste du fonctionnalisme, il faut montrer que les propriétés psychiques sont bel et bien réalisées de diverses façons et que cette réalisation multiple est de nature à empêcher leur réduction. Commençons par la première démonstration, que nous pouvons diviser en deux volets.

Il est raisonnable de croire qu'un humain et un oiseau peuvent partager un état mental, par exemple le désir de manger un certain fruit visuellement aperçu par l'un et l'autre. Mais il est fort peu probable qu'au même moment l'humain et l'oiseau partagent aussi un état neurologique. L'état neurologique occupant le rôle fonctionnel « désirer manger un fruit » dans l'économie psychique de l'oiseau n'est pas similaire à celui qui occupe le même rôle dans l'économie mentale humaine. Bref, l'état psychique « désirer manger un fruit » se trouve réalisé de deux façons distinctes. Puisqu'il n'y a aucune raison de croire que l'état psychique en question est spécial par rapport aux autres états psychiques, et puisqu'il n'y a aucune raison de croire que l'oiseau est un animal spécial sur le plan psychologique par rapport aux autres animaux, on peut généraliser et affirmer que *les états psychiques sont réalisés de manières distinctes chez les animaux de différentes espèces*. Nous appellerons « interspécifique » cette forme de réalisation multiple entre diverses espèces animales. Il existe aussi une forme plus forte de réalisation multiple que nous appellerons « intraspécifique », car elle prévaut au sein d'une même espèce.

Le cerveau des personnes dont l'hydrocéphalie fut corrigée à la naissance est grandement atrophié et, pourtant, ces personnes sont d'intelligence normale et ne manifestent aucune pathologie cognitive particulière. Il est cependant raisonnable de croire que leurs états psychiques ne sont pas réalisés de la même façon que chez les autres humains. Même chez les humains dont le cerveau est anatomiquement normal, il existe sûrement des différences cérébrales importantes, d'un individu à un autre, au niveau de la structure fine, c'est-à-dire dans le nombre de neurones, le nombre de synapses, la localisation et la force des connexions synaptiques, etc. Puisque tous ces individus peuvent néanmoins partager des états psychiques, il est raisonnable de croire que les états psychiques sont en général réalisés de diverses façons chez les différents individus d'une même espèce. Il paraît aussi raisonnable de croire qu'à ce niveau de description fine, la réalisation multiple intraspécifique ne se limite pas aux différents individus mais qu'elle s'étend jusqu'à l'individu particulier. Nous apprenons vers cinq ans que le père Noël n'existe pas. On dit par ailleurs que chaque individu perd environ un millier de neurones quotidiennement, ce qui signifie que notre croyance s'est maintenue malgré la perte de millions de neurones. Il est donc peu probable que cette croyance soit finement réalisée de la même façon chez un même individu, à l'âge de cinq ans et, par exemple, lorsqu'il a plus de vingt ans. Nous avons de plus appris de nombreuses choses

durant toutes ces années, y compris au sujet du père Noël, et ces connaissances ont certainement modifié la structure fine de nos cerveaux. Bref, il semble raisonnable de croire que les états psychiques d'une personne sont réalisés de diverses façons à divers moments de sa vie¹⁰. Comme tous ces cas se généralisent, on peut donc affirmer que les états psychiques sont réalisés de diverses façons chez différents types d'organismes, chez différents organismes d'un même type et chez le même organisme à différents moments. La réalisation multiple des états psychiques semble ainsi radicale.

Pour le fonctionnalisme, le seul déterminant de l'identité d'un état psychique réside dans le rôle fonctionnel qu'il occupe. La thèse de la réalisation multiple affirme simplement que des états de types physiques distincts peuvent occuper le rôle fonctionnel définissant un même état psychique. La réalisation multiple des états psychiques est donc tout à fait compatible avec une conception fonctionnaliste de la nature des états psychiques. Ce fait constitue bien sûr un argument empirique important en faveur du fonctionnalisme. Mais certains fonctionnalistes ont cherché à tirer un argument antiréductionniste à partir de la thèse de la réalisation multiple. Autrement dit, selon ce point de vue, cette dernière fournirait non seulement un argument en faveur du fonctionnalisme, mais également un argument contre le réductionnisme. Ainsi, le fonctionnalisme et le réductionnisme entretiendraient une relation d'exclusion mutuelle relativement à la réalisation multiple. Pour justifier cette dernière position, il faut cependant montrer que la réalisation multiple des états psychiques empêche leur réduction à des états neurologiques. C'est ici, et seulement ici, que le réductionniste doit s'inscrire en faux contre le fonctionnaliste antiréductionniste¹¹.

Complétons d'abord l'argument antiréductionniste fondé sur la réalisation multiple. Remarquons à cet effet que le principe suivant exprime une condition minimale du réductionnisme :

- (1) Toute propriété ne se réduit qu'à une et une seule propriété, ou un et un seul ensemble de propriétés.

Cette condition exprime simplement un aspect du sens du verbe « réduire » tel que l'utilise le réductionniste : si *A* se réduit à *B*, alors il ne peut pas aussi se réduire à *C*. La relation de réduction partage ainsi un trait essentiel avec la relation d'identité. Si *A* est identique à *B* (l'étoile du soir est identique à l'étoile du matin) et si *B* est distinct de *C* (l'étoile du matin est distincte de l'étoile du Nord) alors *A* est aussi distinct de *C* (l'étoile du soir est distincte de l'étoile du Nord). Ce parallèle n'est pas fortuit. La réduction a pour fonction d'élaguer notre ontologie en montrant les identités au sein de l'ensemble des

10. Horgan, T., « Nonreductive Materialism and the Explanatory Autonomy of Psychology », dans Wagner, S. et Wagner, R., dir., *Naturalism : A Critical Appraisal*, Notre Dame, Notre Dame University Press, 1993.

11. Churchland, P. S., *Neurophilosophy*, Cambridge (Mass.), MIT Press, 1986 ; Kim, *Mind in a Physical World*.

propriétés et, pour ce faire, elle doit partager la structure logique de l'identité ; bref, dans l'usage réductionniste contemporain, « réduire » doit impliquer « est identique à » et ne peut le faire qu'à condition de partager cet aspect de la structure logique de l'identité¹². Comment peut-on alors soutenir que le désir de manger un fruit se réalise de deux manières distinctes et que le même désir ne se réduise qu'à un et un seul ensemble de propriétés physiques? La réalisation multiple semble contredire le réductionnisme, dans la mesure où elle entraînerait un conflit entre la condition minimale (1) et des faits comme ceux que nous venons de noter, à savoir :

(2) La réalisation physique du désir que *p* est *X*.

(3) La réalisation physique du désir que *p* est *Y*.

Ainsi, l'incompatibilité entre ces trois énoncés montrerait que la réalisation multiple des états psychiques est incompatible avec leur réduction à des états physiques. C'est du moins ce que prétend l'antiréductionniste. Quelle raison a-t-on de le croire?

2. Pourquoi la conclusion antiréductionniste ne découle pas de l'argument fonctionnaliste

Considérons d'abord la réduction interspécifique ; nous discuterons ensuite la réduction intraspécifique. Dans un compte-rendu de l'ouvrage où Putnam a présenté son argument antiréductionniste, Lewis souligne que le conflit entre la réalisation multiple et le principe minimal du réductionnisme, exemplifié ici par l'incompatibilité entre les énoncés (1)-(3), est de même nature que le « conflit » qui résulterait de l'affirmation que les numéros gagnants à la loterie sont 03 et 61¹³. Puisqu'il n'y a qu'un et un seul numéro gagnant à la loterie, comment 03 et 61 peuvent-ils être tous les deux des numéros gagnants? Lorsque le conflit est posé ainsi, il devient évident qu'il disparaîtra aussitôt qu'on précisera le caractère relatif du prédicat « est un numéro gagnant ». Les loteries tirent de nouveaux numéros gagnants sur une base

12. Nagel, dans *The Structure of Science*, souligne que la théorie classique de la réduction n'implique pas nécessairement l'identité mais, comme le remarque Fodor, dans « Special Sciences », la plupart des philosophes réductionnistes ont opté pour une interprétation *forte* de la réduction, qui implique des identités, car elle seule garantit la généralité de la physique. Une réduction qui n'implique pas des identités est compatible avec un dualisme de propriétés ou d'événements qui sied mal avec l'usage contemporain que les réductionnistes font de la théorie de la réduction. Cette différence d'opinion entre Nagel et Fodor s'explique par les intérêts différents qui motivent ces philosophes. Comme nous l'avons noté au début de cet article, l'objectif de ceux qui ont formulé le modèle classique de la réduction n'était pas d'élaguer l'ontologie, puisque les philosophes n'avaient pas, selon eux, à s'occuper de telles questions métaphysiques, mais d'intégrer les théories de niveau supérieur, comme la psychologie, dans la structure de la science unifiée.

13. Voir Lewis, D., « Review of *Art, Mind, and Religion* », *The Journal of Philosophy*, 66, 1969, p. 23-35.

hebdomadaire, de sorte qu'il y a un nouveau numéro gagnant à chaque semaine : le numéro gagnant de la semaine dernière est (ou était) 03, et celui de cette semaine est 61. De la même manière, soutient Lewis, l'apparente incompatibilité entre la réalisation multiple et la condition minimale du réductionnisme se dissipe dès lors que nous précisons le caractère relatif des prédicats psychologiques :

(1') Toute propriété *d'un type de système* ne se réduit qu'à une et une seule propriété (ou un et un seul ensemble de propriétés) *de ce type de système*.

(2') La réalisation physique du désir que *p dans un système cognitif de type A* est *X*.

(3') La réalisation physique du désir que *p dans un système cognitif de type B* est *Y*.

Le nouveau principe minimal du réductionnisme relativisé à des types de systèmes cognitifs partage toujours la structure logique de l'identité. Si *A-dans-X* est identique à *B-dans-X*, rien n'empêche que *A-dans-Y* soit distinct de *B-dans-X*. Le principe permet ainsi d'élaguer l'ontologie tout en respectant la réalisation multiple des prédicats psychologiques. Mais il faut souligner que la relativisation des prédicats psychologiques ne limitera leur extension d'une manière qui convient à la construction de lois psychologiques et de règles de correspondance que dans la mesure où les types de systèmes cognitifs concernés sont ou font partie d'espèces naturelles (*natural kinds*). Sinon, l'antiréductionniste pourra objecter, avec raison, que la relativisation n'est qu'une astuce *ad hoc* servant à prémunir le réductionnisme contre l'objection de réalisation multiple.

Il est assez naturel de laisser à la biologie le soin d'identifier les types de systèmes cognitifs qui pourront servir à la relativisation de la portée des prédicats psychologiques. La sélection naturelle a « construit » les systèmes cognitifs et les a regroupés en types. On peut dès lors penser qu'elle a produit des différences que voudra saisir la psychologie, ou du moins des différences dont celle-ci acceptera de s'accommoder. L'autonomie absolue de la psychologie par rapport aux sciences limitrophes n'est plus une position valorisée, tant en psychologie qu'en philosophie de l'esprit ou des sciences. Suivant cette piste, plusieurs philosophes ont proposé de limiter l'extension des prédicats psychologiques aux espèces biologiques¹⁴. Les prédicats psychologiques ne seraient pas transférables d'une espèce à une autre, même si l'on parle habituellement comme s'ils l'étaient. Ainsi, bien qu'il existe un usage populaire permettant l'attribution du prédicat « croire que *p* » aux humains et aux bêtes, il est raisonnable de postuler qu'un discours bien policé sur le plan scientifique interdira cet usage permissif du prédicat. Mais s'il s'avère

14. Voir, par exemple, Mundale, J. et Bechtel, W., « Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function », *Minds and Machines*, 6, 1996, p. 481-505.

raisonnable de relativiser ainsi les prédicats d'attitudes propositionnelles — en supposant que nous admettions, avec l'antiréductionniste, qu'ils feront partie d'une psychologie scientifique future —, il semble tout aussi raisonnable de rejeter la relativisation d'autres prédicats psychologiques, notamment ceux initialement considérés par Putnam, soit la faim et la douleur. Le réductionniste ne disposerait donc pas des ressources nécessaires pour saisir plusieurs des prédicats fondamentaux de la psychologie.

Mais ce serait là mésestimer les ressources de la biologie. L'ensemble des espèces animales ne forme pas une classe mais une structure liée par des relations de parenté, c'est-à-dire un arbre dont les nœuds sont des espèces et les arêtes des relations de descendance. Puisque le processus à l'origine de ces liens ne « crée » pas les organismes de toutes pièces mais, en général, modifie légèrement l'espèce-mère, la parenté biologique des espèces reflétera une parenté structurelle et fonctionnelle qui aura une incidence sur les sciences prenant ces espèces ou leurs propriétés comme objet. On peut croire, par conséquent, que les prédicats psychologiques se relativiseront à différentes structures évolutionnistes et qu'il existera un rapport entre le degré de parenté entre deux espèces et leur degré de similarité psychologique¹⁵. Comme le notent Bechtel et McCauley, les espèces apparentées partagent des structures cérébrales et il est raisonnable de faire l'hypothèse que ces structures ont des fonctionnalités identiques ou apparentées (*Heuristic Identity Theory*)¹⁶. Certaines propriétés psychologiques pourront ainsi se relativiser à l'espèce, d'autres au genre, d'autres à la famille, d'autres à l'ordre, et ainsi de suite¹⁷. Pour reprendre l'exemple initial de Putnam, on peut penser que la faim se relativise à un très vaste ensemble d'espèces, peut-être à tout le règne *Metazoa*, la nutrition étant une des caractéristiques qui définit le règne animal. En fait, si l'on considère le mode par lequel la sélection naturelle produit de nouvelles espèces, on peut croire que seul un petit sous-ensemble des prédicats psychologiques se relativisera à l'espèce. Croit-on vraiment que chacune des quatorze espèces de pinsons identifiées par Darwin sur l'archipel de Galapagos manifeste des propriétés psychologiques distinctes? Dans ce cas particulier, il sera sans doute préférable de relativiser les prédicats psychologiques à la famille (*Fringillidae*). L'espèce humaine est peut-être l'exception qui confirme la règle puisque la différence anatomique principale qui la distingue des ses plus proches parents (tant ancestraux que contemporains) résulte d'un processus d'encéphalisation qui a multiplié la taille du néocortex. L'espèce humaine se distinguera ainsi par la quantité de propriétés

15. La proposition inverse ne tient pas : la similarité psychologique n'entraîne pas nécessairement une parenté biologique. On parlerait plutôt dans ce cas d'homologies, comme les ailes des oiseaux, des insectes et des chauves-souris.

16. Voir Bechtel et McCauley, « Heuristic Identity Theory (or Back to the Future) : The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience ».

17. Si on utilise la conception cladistique de la classification des organismes, on pourra relativiser les prédicats au niveau du clade, un clade étant un groupe d'embranchements d'animaux ayant une même organisation et une évolution phylétique commune.

psychologiques qui lui sont uniques et sera, par conséquent, celle possédant le plus grand nombre de propriétés relativisées à l'espèce.

L'antiréductionniste objectera peut-être que le prédicat « est un numéro gagnant » a toujours été relatif à un tirage particulier alors que le prédicat « désirer que p » n'est pas relatif à une espèce donnée, tant au sein de notre psychologie populaire que de notre psychologie scientifique actuelle. Une telle objection trahit toutefois un penchant pour une conception inacceptable de la science, ou pour une forme de dualisme, avec laquelle l'antiréductionniste ne voudra sûrement pas être associé. L'objection présuppose que les prédicats utilisés dans les diverses sciences sont immuables : s'ils n'étaient pas relatifs dans une théorie T , alors ils ne pourront pas l'être dans toute théorie postérieure à T qui entend traiter du même domaine. Ce principe, que nous pourrions appeler « le principe de l'immutabilité absolue des prédicats scientifiques », fait cependant de notre antiréductionniste un homme de paille : ce principe n'a jamais été accepté en sciences puisqu'il entrave considérablement son développement. Où serait la physique actuelle si les philosophes avaient réussi à imposer le principe de l'immutabilité absolue des prédicats scientifiques pour empêcher la réforme des concepts de notre physique naïve¹⁸? Comme le remarque Thagard¹⁹, la science progresse en grande partie par des « révolutions conceptuelles », c'est-à-dire par l'élagage des concepts superflus, la création de nouveaux concepts (soit *de novo*, soit par la division ou la redéfinition de l'extension d'anciens concepts), et enfin la réorganisation des relations genre-espèce et partie-tout entre les concepts.

Nous savons par exemple qu'en dépit de ce que nous enseigne notre physique naïve, il s'avère que la propriété macroscopique de température, disons T_{macro} , se trouve réalisée par de nombreuses propriétés microscopiques distinctes T_{micro1} , T_{micro2} , ..., $T_{micro \cdot n}$. Pour un humain, rien ne ressemble plus à la chaleur de l'air que celle de l'eau et pourtant il s'avère que la réalisation de la température est relative aux états de la matière. La température d'un gaz est l'énergie cinétique moyenne des molécules du gaz, celle d'un solide est l'énergie cinétique maximale des molécules du solide, et celle d'un plasma ne peut être définie en termes d'énergie moléculaire puisqu'un plasma ne contient pas de molécules, mais seulement des atomes ionisés. Le terme général « température » dénote donc plusieurs propriétés macroscopiques distinctes (température d'un gaz, d'un liquide, etc.) qui partagent une certaine similarité superficielle, correspondant chacune à des propriétés microscopiques distinctes. Cela signifie-t-il que la température est irréductible? Au contraire : la réduction de la température est un cas paradigmatique de réduction justifiée²⁰. S'il a été possible de réduire la température et ainsi

18. McCloskey, M., « Intuitive Physics », *Scientific American*, 248, 1983, p. 122-130.

19. Thagard, P., *Conceptual Revolutions*, Princeton, Princeton University Press, 1992.

20. Voir Enç, B., « In Defense of the Identity Theory », *The Journal of Philosophy*, 80, 1983, p. 279-298.

de redéfinir le concept de température, c'est parce que l'extension originelle du concept était bien structurée et qu'il a dès lors été possible de redéfinir le concept d'une manière non *ad hoc* permettant la formulation de lois scientifiques. L'extension de T_{macro} peut être ainsi divisée sans reste en cinq sous-ensembles disjoints, et il est raisonnable de croire que toute nouvelle exemplification pourra se grouper sous l'un ou l'autre de ces sous-ensembles sauf, bien sûr, si l'on découvrait un jour un nouvel état de la matière.

Les états de la matière permettent de relativiser naturellement la propriété macroscopique, c'est-à-dire de la découper d'une manière qui respecte la structure fine de la nature. De la même manière, les espèces animales, ou même extraterrestres et robotiques, permettent de relativiser les propriétés psychologiques d'une manière qui respecte aussi la structure fine de la nature. Ici aussi, l'extension du prédicat originel est bien structurée car elle peut être divisée en sous-ensembles disjoints, et il est raisonnable de croire que toute nouvelle exemplification pourra se grouper sous l'un ou l'autre de ces ensembles — sauf si l'on découvre une nouvelle espèce, famille, etc., le cas des extraterrestres ou des robots n'étant pas différent de celui des états encore inconnus de la matière. Les nouvelles extensions créées par cette partition de l'extension du prédicat originel permettront la construction des nouveaux prédicats macroscopiques relativisés qui se réduiront, un à un, à des propriétés microscopiques déterminées.

Le réductionniste peut donc répondre à l'argument de la réalisation multiple *interspécifique* en deux étapes : il montre d'abord que celle-ci demande une modification légitime des prédicats de la psychologie, soit leur relativisation à des espèces, familles, etc., qui permet toujours la formulation de lois psychologiques relativisées ; il rappelle ensuite que la réforme conceptuelle est une étape essentielle du développement de toute science, y compris la psychologie. Une réponse aussi simple, rapide (elle date de 1969, soit deux ans après la parution de l'argument) et décisive aurait dû faire oublier assez rapidement cet argument. Mais force est d'admettre qu'il est toujours aussi populaire trente ans plus tard. Je soupçonne que sa ténacité s'explique ainsi²¹ : la réalisation multiple interspécifique est un outil rhétorique servant à introduire une autre forme de réalisation multiple qui, elle, empêche bel et bien la formulation de lois de correspondance. Elle empêche la construction de lois de correspondance en niant l'existence de lois psychologiques générales. Les lois psychologiques doivent en effet subsumer des individus psychologiques mais, si cette autre forme de réalisation multiple s'avère vraie, chaque loi psychologique ne pourra subsumer qu'un et un seul individu (voire un individu à un et un seul moment de son existence). Une telle restriction de l'extension des lois et des

21. Je ne m'attarderai pas ici aux facteurs idéologiques qui faisaient à l'époque, et encore aujourd'hui, de l'antiréductionnisme une position de bon goût ou, comme nous dirions maintenant, politiquement correcte. S'avouer réductionniste en matière psychologique dans le milieu académique contemporain, c'est un peu comme s'avouer un penchant pour la droite, la musique disco ou la viande rouge.

prédicats psychologiques, bien qu'elle soit possible, redéfinirait la psychologie d'une manière inacceptable. Nous verrons cependant que nous pouvons rejeter cette autre forme de réalisation multiple.

Notre antiréductionniste pourrait enfin être tenté de riposter que certaines caractéristiques de la psychologie font que le principe de l'immutabilité des prédicats s'y applique. Pensons, par exemple, au caractère intime de notre rapport à nos états psychiques ou à l'incapacité de connaître le contenu de l'esprit d'autrui. Par cette réponse, l'antiréductionniste s'engage cependant à un dualisme qui postule d'entrée de jeu que l'appareil méthodologique de la science ne s'applique pas en psychologie. Un mécanisme fondamental du développement de toute science a pour tâche de préciser l'extension des prédicats qu'elle utilise à la lumière de nouvelles données ou de nouvelles théories, et quiconque nie, pour les raisons que nous avons notées ou d'autres, l'action de ce mécanisme en psychologie exclut d'entrée de jeu cette discipline du champ des sciences. Il n'est pas impossible, bien sûr, que la psychologie ne soit pas une science, ou pas une science comme les autres, et que le principe de l'immutabilité des principes s'y applique en particulier. Mais on ne peut adopter cette position qu'après avoir laissé agir le mécanisme de réforme conceptuelle et non avant son action, pour nier son applicabilité. Je ne m'attarderai pas sur cette riposte, bien qu'elle soit populaire. Elle fait reposer l'antiréductionnisme sur un antiphysicalisme, et je présuppose que l'antiréductionniste qui nous intéresse ici insiste pour conserver à la psychologie son caractère scientifique. Bref, nous ne sommes pas concernés ici par celui qui achète son antiréductionnisme au moyen d'un chèque antiphysicaliste. Considérons maintenant la réalisation intraspécifique.

La relativisation des prédicats psychologiques aux espèces ou autres groupes biologiques identifiés par leurs propriétés phylogénétiques bloque l'inférence allant de la réalisation multiple *interspécifique* à l'irréductibilité du psychique. Les espèces, ou autres groupes biologiques, permettent une partition empiriquement motivée de l'extension des prédicats psychologiques et il semble qu'il soit tout à fait justifié, sur le plan scientifique, de restreindre les lois psychologiques aux propriétés ainsi relativisées. Mais s'il est raisonnable de limiter les lois psychologiques aux divers groupes biologiques, il n'est pas raisonnable de les limiter aux individus de l'espèce. Le réductionniste peut répondre à l'antiréductionniste que, bien qu'elle soit logiquement possible, sa conception de la psychologie comme science générale dont *toutes* les lois doivent en principe pouvoir s'appliquer tant aux pieuvres qu'aux humains et aux Martiens n'est pas la psychologie scientifique qui se pratique aujourd'hui (ou qu'elle n'en constitue qu'une infime partie²²). Et il peut aussi objecter que la relativisation des prédicats psychologiques aux groupes biologiques découpe mieux la nature qu'une division incluant les pieuvres, les humains et les Martiens sous une même catégorie. Mais il ne peut du même coup endosser une

22. Mundale et Bechtel, « Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function ».

psychologie individualiste qui, bien qu'elle soit elle aussi logiquement possible, ne correspond pas à la psychologie telle qu'elle se pratique aujourd'hui. Et il ne peut pas non plus demander à la psychologie de demeurer aveugle aux lois interindividuelles qui existent manifestement en psychologie. Bref, le réductionniste ne peut répondre à la réalisation multiple *intraspécifique* comme il l'a fait pour la réalisation *interspécifique*, soit en relativisant les prédicats à des groupes plus petits que l'ensemble des systèmes cognitifs possibles. Dans ce qui suit, je montrerai comment on peut rejeter l'inférence de l'irréductibilité du domaine psychique au domaine physique à partir de la réalisation multiple *intraspécifique*.

Revenons à la réduction de la température. Nous avons déjà vu que la relativisation de la température aux divers états de la matière suggère une analogie intéressante avec la relativisation des prédicats psychologiques aux diverses espèces, ce qui permet de répondre au problème posé par la réduction interspécifique. Or, comme le note Bickle²³, la réduction de la température suggère également une analogie pertinente pour le problème posé par la réalisation multiple *intraspécifique*. À un niveau de description neurologique suffisamment fin, il y a réalisation multiple de la croyance en l'inexistence du père Noël, puisque celle-ci se maintient malgré la perte de millions de neurones. Mais à un niveau de description suffisamment fin, deux températures identiques dans un même gaz seront presque toujours réalisées de manières distinctes, et cela par nécessité nomologique²⁴. Lorsque la description physique est suffisamment fine pour saisir les différences individuelles entre deux volumes d'un même gaz, notamment la vitesse et la masse des molécules qu'ils contiennent, deux températures identiques seront presque nécessairement réalisées de manières distinctes. On a donc là, du moins en surface, un cas analogue à celui de la réalisation multiple *intraspécifique* des propriétés psychologiques. Voyons pourquoi la réalisation multiple *intraspécifique* de la température ne perturbe pas la thermodynamique des gaz ou la réduction classique de la température, et si nous pouvons en tirer une leçon pour la psychologie.

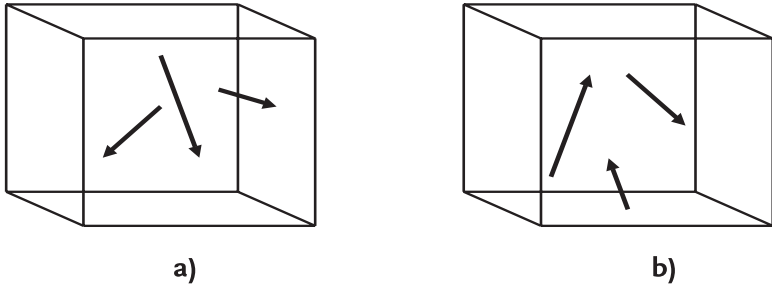
Les cubes (a) et (b) ci-dessous représentent des volumes identiques d'oxygène contenant seulement trois molécules d'O₂²⁵. Les vecteurs représentent les vitesses respectives des molécules. Au niveau de description représenté par le schéma, les deux gaz présentent des propriétés microscopiques distinctes.

23. Bickle, *Psychoneural Reduction*.

24. Les chances que deux températures identiques soient finement réalisées de manière identique sont astronomiquement petites. Cela signifierait que les deux volumes de gaz contiennent le même nombre du même type de molécules ayant toutes la même vitesse (vitesse et direction). Étant donné qu'un mètre cube d'air contient environ 10²⁵ particules, il est raisonnable de croire que cette situation ne s'est pas produite depuis le début de l'univers.

25. Ce schéma est de Bickle, *Psychoneural Reduction*, p. 125.

Figure 1



La température d'un gaz est proportionnelle à l'énergie cinétique moyenne de ses molécules. La température des deux volumes sera par conséquent identique si l'équation suivante est valide :

$$\frac{\sum_n^1 \frac{1}{2} m_a v_a^2}{n} = \frac{\sum_n^1 \frac{1}{2} m_b v_b^2}{n} \quad (1)$$

Puisque nos deux volumes (a) et (b) contiennent un nombre identique de molécules de masse identique (la masse de deux molécules d'un même type est toujours identique : 32 000 *amu* dans le cas d'une molécule de O_2), l'équation se réduit ainsi :

$$v_{a1} + v_{a2} + v_{a3} = v_{b1} + v_{b2} + v_{b3} \quad (2)$$

Puisque, par hypothèse, seule la composante directionnelle de la vélocité des molécules diffère (les trois vecteurs étant de même longueur), et que cette composante est négligée par la formule (il n'y a pas de facteur la représentant), l'équation se réduit encore :

$$0 = 0 \quad (3)$$

Nos deux températures sont donc bel et bien identiques mais réalisées de manière distinctes, car les vélocités des molécules diffèrent. Prenons maintenant un seul volume de gaz dont la température demeure constante pour un certain temps. Puisqu'il y aura des collisions et par conséquent des transferts d'énergie cinétique, et puisque les molécules se déplacent (elles ont une énergie cinétique), il s'ensuit que la température identique sera réalisée de manière distincte à n'importe lesquels de deux moments de cette période de temps. Ce cas est plus intéressant que le premier, puisque les collisions amènent des changements de vitesses, c'est-à-dire des changements dans un facteur considéré par l'équation. Mais ces changements de vitesse s'annulent globalement, puisque la vitesse perdue par une molécule sera gagnée par une

autre. Une mesure statistique qui fait la somme des vitesses est donc aveugle à ce genre de transfert énergétique.

Si ces cas de réalisation multiple ne troublent pas les partisans de la réduction thermodynamique de la température, c'est-à-dire l'ensemble de la communauté scientifique contemporaine, c'est que la réduction de la température « ne descend pas²⁶ » vraiment jusqu'au niveau des propriétés moléculaires des gaz. Elle s'arrête plutôt à un niveau un peu plus élevé où certaines propriétés moléculaires, comme la composante directionnelle de la vélocité, la vibration et la rotation des molécules, etc., ne sont pas considérées. La composante directionnelle est ignorée car, dans un ensemble suffisamment grand, on peut faire l'hypothèse que toutes les directions seront représentées et donc que la direction globale des molécules sera nulle. Mais si les molécules avaient une direction préférentielle, si elles montaient par exemple, il y aurait alors diminution de la pression dans le bas du volume et, donc, diminution concomitante de la température à cet endroit — c'est d'ailleurs ce qui se produit dans l'atmosphère lorsqu'un front froid se déplace. La vibration et la rotation des molécules sont aussi des mouvements moléculaires qui contribuent à la température réelle du gaz. Celles-ci sont toutefois négligées par la formule thermodynamique classique car elles ont un effet négligeable sur la température du gaz. Elles ont cependant leur importance, par exemple, en cryogénie ou lorsqu'on cherche à s'approcher du zéro absolu. La communauté accepte la réduction de la température mais sait pertinemment que celle-ci représente une idéalisation par élimination de facteurs ayant une influence bien réelle mais négligeable. À ce même niveau de description, d'autres propriétés moléculaires ne sont pas négligées mais prises globalement, par exemple la vitesse et la masse, qui sont sommées. C'est pourquoi les changements de vitesse résultant de collisions élastiques (où il n'y a pas perte d'énergie cinétique) s'annulent, une molécule acquérant exactement la vitesse perdue par l'autre. Enfin, à ce même niveau de description, d'autres propriétés encore ne sont considérées que dans une forme idéalisée. Le volume des molécules est postulé être nul (particules ponctuelles), ce qui n'est évidemment pas le cas en réalité. Les collisions, comme nous l'avons déjà remarqué, sont postulées être parfaitement élastiques, etc.

La réduction est établie entre la propriété macroscopique, en l'occurrence la température du gaz, et une propriété introduite à un niveau *abstrait* et *idéalisé* de description mathématique du comportement global des molécules du gaz, soit l'énergie cinétique moléculaire moyenne. Un mètre cubique d'air contient environ 10^{25} particules²⁷. Bien que celles-ci se comportent comme un système newtonien et que le concept central de la réduction (l'énergie cinétique) soit un concept mécanique classique (c'est une conséquence directe de la seconde loi de Newton), il est impossible de calculer les interactions au niveau des molécules individuelles. Il est par conséquent

26. Nous poursuivons ici la métaphore de la verticalité inhérente au concept de réduction.

27. Concevoir des molécules comme des particules constitue aussi une idéalisation.

impossible de traiter le gaz comme un système newtonien : il est difficile de calculer les états d'un système newtonien comprenant trois masses comme, par exemple, le soleil, la terre et la lune — ce qui en dit long sur la possibilité de faire le même calcul pour 10^{25} masses. La mécanique statistique et la thermodynamique statistique furent créées pour résoudre ce problème²⁸.

Bref, la réduction de la température ne se rend pas jusqu'au niveau des propriétés microscopiques des molécules du gaz mais s'arrête au niveau d'une description mathématique plus abstraite et idéalisée. Lorsqu'on descend à un niveau de description plus fin, on observera une réalisation multiple intraspécifique. Ces remarques au sujet de la réduction de la température sont relativement évidentes. Elles montrent qu'on pourra toujours générer une accusation de réalisation multiple intraspécifique si l'on choisit des propriétés suffisamment microscopiques par rapport à la propriété macroscopique dont on entend nier la possibilité de réduction. Elles montrent aussi comment le réductionniste doit répondre à ces accusations : montrer qu'il existe, ou peut exister, un niveau de description *microscopique* où il n'y a pas de réalisation multiple intraspécifique. Car il n'est pas suffisant de remarquer l'illégitimité de l'attaque. Encore faut-il montrer qu'il est possible de définir un niveau de description qui permette de décrire les propriétés des neurones au bon niveau d'abstraction. L'antiréductionniste pourrait en effet répondre : « Merci pour la leçon de physique 101. Mais la neurologie diffère de la physique en ceci : le niveau de description où se pose la réalisation multiple intraspécifique est le seul niveau de description offert par la neurologie, et le seul possible ».

Le débat se réduit donc à la question de savoir s'il existe un équivalent neurologique du niveau intermédiaire de description de la thermodynamique statistique, un niveau de description qui permette de traiter des propriétés des neurones et des circuits neuronaux d'une manière globale, faisant abstraction des détails inutiles ou gênants. Il faut par exemple un niveau de description qui fasse abstraction du nombre de neurones puisqu'une croyance peut se maintenir malgré la perte quotidienne de neurones et puisque deux individus peuvent partager des croyances sans partager le même nombre de neurones. Il faut aussi un niveau de description qui fasse abstraction du détail des interconnexions synaptiques — quel neurone est connecté à quel autre, et avec quelle force — puisqu'une croyance peut se maintenir malgré la formation, la perte et la variation dans la force des interconnexions et puisque deux individus peuvent partager des croyances sans partager une matrice d'interconnexions. Il faut aussi un niveau de description qui fasse abstraction des différences d'enco-

28. Il n'est sans doute pas inutile dans le cadre du débat actuel au sujet de la réduction des propriétés psychologiques de rappeler que ces idées ne furent pas immédiatement acceptées en physique. Plusieurs physiciens voyaient alors d'un mauvais œil et accusaient de « mécaniste » toute tentative de réduire les propriétés électromagnétiques ou thermodynamiques à la mécanique classique (où les corps impliqués seraient des atomes). Je rappelle ces faits pour contrer notre tendance à surestimer les différences entre la physique et les autres sciences. En physique, comme ailleurs, la réduction n'est ni donnée, ni automatique, ni facile.

dage de l'*input* puisque deux individus peuvent partager des croyances au sujet d'un objet malgré la provenance distincte des informations initiales (visuelle et tactile, par exemple). De toute évidence, ce niveau de description devra pouvoir faire abstraction d'un ensemble d'autres différences microscopiques n'ayant pas de répercussion au niveau macroscopique, mais concentrons-nous sur celles que nous venons de nommer.

On pourrait croire que ce niveau de description nous viendra des neurosciences mathématiques (*computational neuroscience*). Les réseaux de neurones artificiels constituent des modèles intéressants, car ils se situent à un niveau intermédiaire entre la psychologie et la neurologie. Ils possèdent à la fois des propriétés microscopiques *semblables* à celles de réels systèmes nerveux et des propriétés macroscopiques *semblables* à celles de réels sujets psychologiques. Par exemple, des systèmes constitués d'unités simples de calcul modélisant certaines des propriétés que l'on retrouve dans les assemblées nerveuses parviennent à départager des énoncés grammaticaux et non grammaticaux²⁹. Ces propriétés des réseaux de neurones devraient à tout le moins les rendre intéressants aux yeux de ceux qui s'intéressent à la réduction psychophysique. Comme la thermodynamique statistique, ils semblent offrir un niveau de description intermédiaire ouvrant la possibilité de réduire les propriétés psychologiques aux propriétés d'un système abstrait et partiellement idéalisé qui, sous les conditions appropriées, se réduisent elles-mêmes aux propriétés microscopiques visées. Puisque la réduction est transitive, cette réduction psychophysique en deux étapes constituerait une réduction psychophysique en bonne et due forme, tirant profit d'un pont entre les propriétés macroscopiques de la psychologie et celles, trop microscopiques, de la neurologie.

Mais un réseau de neurones est un système relativement simple, entièrement décrit au niveau microscopique par sa fonction d'activation, son biais, son schème d'interconnexion et sa matrice de coefficients synaptiques³⁰. Deux réseaux de neurones possédant la même fonction d'activation, le même biais, le même schème d'interconnexion et la même matrice de coefficients synaptiques sont identiques autant sur le plan microscopique que macroscopique. Cependant, deux réseaux initialement identiques au niveau microscopique peuvent être entraînés à réaliser la même tâche avec un taux de succès identique, c'est-à-dire posséder exactement la même propriété macroscopique et, pourtant, selon le corpus d'entraînement utilisé, posséder des matrices de

29. Voir Elman, J. L., « Learning and Development in Neural Networks: The Importance of Starting Small », *Cognition*, 48, 1993, p. 71-99 ; Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., et Plunkett, K., *Rethinking Innateness*, Cambridge (Mass.), MIT Press, 1996.

30. Je ne veux pas présenter ici une introduction générale aux réseaux de neurones artificiels. Le lecteur qui ne serait pas familier avec ces formalismes pourra lire l'excellente introduction de W. Bechtel et A. A. Abrahamsen, *Le connexionnisme et l'esprit : introduction au traitement parallèle par réseaux*, traduction de J. Proust, Paris, La Découverte, 1993.

coefficients synaptiques distinctes. Deux réseaux de neurones artificiels identiques au niveau macroscopique peuvent donc différer au niveau microscopique. Les propriétés macroscopiques des réseaux de neurones se réalisent de diverses façons. Mais, bien qu'il soit abstrait et idéalisé par rapport au niveau de description propre à la neurologie, le niveau microscopique de description propre aux réseaux de neurones artificiels ne peut remplir le rôle d'intermédiaire puisque ces systèmes reproduisent la relation macro/micro que l'on retrouve entre la psychologie et la neurologie. Ce n'est donc pas à ce niveau qu'on trouvera avantage à utiliser ces modèles artificiels.

Les neurosciences mathématiques offrent toutefois la possibilité de définir mathématiquement les propriétés macroscopiques (« psychologiques ») et microscopiques (« neurologiques ») de réseaux de neurones artificiels. Il est par conséquent possible de leur appliquer l'ensemble des techniques mathématiques d'analyse (analyse des tendances centrales, de la variabilité, factorielle, combinatoire, calcul des probabilités, etc.). Ces techniques permettent de suivre à la trace comment les propriétés microscopiques s'organisent pour permettre au système de manifester ses propriétés macroscopiques ; d'établir des métriques variées entre les différents états microscopiques du système et entre ses différents états macroscopiques ; de déterminer si le réseau possède des propriétés plus abstraites, par exemple si tel neurone a tendance à s'activer avec tel autre, etc. Les analyses mathématiques et les possibilités qu'elles ouvrent ne sont pas absentes en psychologie et en neurosciences mais elles sont beaucoup plus difficiles à définir. En neurosciences mathématiques, les propriétés macroscopiques et microscopiques des systèmes sont avant tout des propriétés mathématiques — les systèmes eux-mêmes sont des formalismes — auxquelles on assigne une interprétation fonctionnelle psychologique (dans le cas des propriétés macroscopiques), ou biologique (dans le cas des propriétés microscopiques)³¹.

Posons qu'un réseau a été entraîné à discriminer deux types d'objets, c'est-à-dire à répondre de manière différenciée aux objets des deux types, par exemple des échos sonars de mines et de rochers³². On pourra regrouper l'activation des divers neurones de la couche cachée du réseau entraîné en un vecteur et représenter celui-ci comme un point dans un espace des configurations³³. Si l'on reporte l'ensemble des vecteurs d'activation produits par les

31. Je ne chercherai pas à évaluer ici la pertinence des interprétations fonctionnelles psychologiques et biologiques ; voir Poirier, P., « Du Stimulus à la Science, neuro-computationnellement », *Actes du XXVII^e Congrès de l'A.S.P.L.F.*, 1999.

32. Voir Gorman, R. P. et Sejnowski, T. J., « Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets », *Neural Networks*, 1, 1988, p. 75-89.

33. Cet espace comprendra une dimension par neurone caché et l'activation du neurone sera représentée par un point dans la dimension appropriée. Ces espaces de configurations sont appelés « espace d'activation » (*activation space*). Ils sont à distinguer des « espaces d'erreur » (*error space*) qui servent à corrélérer l'erreur en output avec l'état synaptique du système à un moment de son entraînement.

différents objets comme autant de points dans un espace d'activation, on remarquera que les objets initiant une réponse particulière tomberont tous dans une certaine région de l'espace et que plus la réponse est confiante — près de 1, si le spectre de réponse des unités de sortie varie de 0 à 1 — plus le vecteur d'activation tombera près d'un point précis de l'espace. L'entraînement aura amené le réseau à créer une partition de son espace d'activation et à placer un « attracteur » au centre de chacune des deux régions. Plus un input causera une activation des neurones rapprochée de l'attracteur, plus la réponse sera claire et confiante. À l'inverse, plus l'input causera une activation des neurones rapprochée de la frontière de la partition, plus la réponse sera ambiguë et indécise.

Ces différentes propriétés, « posséder une partition », « posséder un attracteur », « tomber dans tel ou tel sous-volume de l'espace », sont des propriétés globales du système. Ce sont aussi des propriétés fonctionnelles générales des réseaux entraînés. Un réseau parviendra à accomplir une tâche s'il réussit à ajuster les coefficients synaptiques de sa matrice de manière à établir une partition appropriée de l'espace d'activation. Notre exemple simple ne demandait qu'une bipartition de l'espace, mais une tâche plus complexe demandera la création d'une partition d'une complexité appropriée. Et il s'avère que les limites « cognitives » d'un système correspondent exactement aux partitions qu'il est incapable de construire, soit parce qu'il ne possède pas suffisamment de neurones cachés, et donc que son espace d'activation est trop petit, soit parce qu'une telle partition n'existe pas, soit parce que le corpus d'entraînement ne contient pas les données appropriées, etc. L'ensemble des propriétés macroscopiques qu'un réseau donné peut manifester est donc déterminé par l'ensemble des partitions de l'espace d'activation qu'il peut créer. La création d'une partition de l'espace d'activation est une propriété générale des réseaux de neurones entraînés pour manifester des propriétés macroscopiques. Or, il est apparu récemment que l'espace d'activation de réseaux distincts sur le plan microscopique mais identiques sur plan macroscopique partagent un ensemble de propriétés géométriques. Ce niveau de description, emprunté à la dynamique, et qui parle d'espace de configuration, de partition de l'espace, de sous-volumes, d'attracteurs, de distance entre attracteurs, de parcours dans l'espace, etc., permet la description de propriétés neurologiques globales à un niveau qui fait abstraction des détails de l'implémentation ou de l'histoire des systèmes particuliers.

Dans une série d'expériences, Aarre Laaksso et Garrison Cottrell³⁴ ont entraîné dix réseaux de neurones à discriminer des couleurs. Il s'agit là d'un apprentissage très simple, qui n'est qu'une variation de la tâche de reconnaissance d'échos sonars que nous avons notée ci-dessus. L'objet de l'expérience

34. Laaksso, A. et Cottrell, G., « How Can I Know What You Think? : Assessing Representational Similarity in Neural Systems », dans *Proceedings of the Twentieth Annual Cognitive Science Conference*, Mahwah, Lawrence Erlbaum, 1998.

n'était pas de déterminer si les réseaux de neurones peuvent faire ce type d'apprentissage, puisque nous savons depuis longtemps que ceux-ci sont particulièrement habiles à ce genre de tâches discriminatoires. Le réseau devait simplement apprendre à associer des représentations binaires de noms de couleurs (rouge = 10000, jaune = 01000, etc.) à des données provenant d'un spectrophotomètre analysant un échantillonnage standardisé de couleurs³⁵. Par exemple, le réseau devait associer le nom « rouge » (c'est-à-dire, 10000) à une donnée provenant d'un spectrophotomètre analysant un échantillon de rouge. Nous avons vu que, d'une manière générale, un réseau réussira cette tâche si sa procédure d'apprentissage parvient à diviser l'espace d'activation en cinq sous-volumes tels qu'un input de rouge amène le réseau à produire le nom « rouge », et ainsi de suite. Les dix réseaux entraînés ne variaient que sous un aspect : le nombre de neurones constituant leur couche cachée. Le premier réseau possédait un neurone caché, le second deux, le troisième trois, et ainsi de suite. Le but de l'expérience était de déterminer si les espaces d'activation de ces dix réseaux entraînés à discriminer des couleurs partageaient des propriétés géométriques globales. L'expérience analysait donc la variation des propriétés géométriques de l'espace (variable dépendante) en fonction de la variation du nombre de neurones cachés disponibles pour accomplir la tâche de discrimination (variable indépendante). Avant de passer aux résultats, précisons la pertinence de cette expérience pour la question qui nous occupe. Soulignons d'abord que les dix réseaux possèdent des propriétés microscopiques distinctes : bien qu'ils possèdent les mêmes couches d'entrée et de sortie, leur couche cachée ne partage pas le même nombre de neurones, et il s'ensuit que le nombre de connexions où distribuer les « connaissances » acquises lors de l'apprentissage varie — par exemple, le réseau à 10 neurones cachés possède 119 connexions de plus que celui à trois neurones cachés, chacune de ces connexions ayant un coefficient synaptique spécifique. La variation du nombre d'interconnexions et de coefficients synaptiques est donc de deux ordres de magnitudes. Nous cherchons ici à déterminer si le réductionniste peut comprendre la réalisation multiple intraspécifique des propriétés psychologiques comme on le fait pour la réalisation multiple intraspécifique des gaz, c'est-à-dire en identifiant un niveau de propriétés qui, bien qu'il soit toujours microscopique, soit suffisamment abstrait et global pour que les différences de réalisations perdent leur pertinence théorique. En permettant la description du comportement global d'ensemble de particules, la mécanique statistique réussit ce tour de force dans le cas de la température. Nous voulons savoir si le vocabulaire de la dynamique des systèmes permet une description similaire des propriétés globales d'un réseau de neurones qui fasse abstraction de certains détails d'implémentation comme le nombre de neurones et le détail précis de leur

35. Anonyme, *Kuopio Color Database*, Lappeeranta University of Technology, 1995 ; http://www.lut.fi/ltk/tite/research/color/lutcs_database.html.

interconnexion³⁶. L'expérience de Laaksso et Cottrell observe justement le comportement de l'espace d'activation en fonction de la variation des propriétés microscopiques du réseau (le nombre de neurones cachés et, *a fortiori*, le détail de l'interconnexion).

Il est apparu au cours de l'expérience que la procédure d'apprentissage ne peut pas diviser de manière appropriée l'espace d'activation des réseaux ne comprenant qu'un ou deux neurones cachés. L'espace d'activation de ces réseaux n'est pas suffisamment « spacieux » pour accommoder les sous-espaces nécessaires à la réalisation de la tâche et, par conséquent, ces réseaux n'ont pas réussi à discriminer les cinq couleurs. Tous les autres réseaux ont su créer une partition appropriée et donc discriminer les couleurs. Ils furent entraînés jusqu'à ce qu'ils réussissent la tâche avec un taux de succès similaire. Pour chacun des huit réseaux capables de discriminer les couleurs, Laaksso et Cottrell ont mesuré la distance euclidienne entre le point causé par un input donné (disons, une représentation d'une donnée spectrophotométrique provenant de l'analyse d'un échantillon de rouge) et les points causés par les autres inputs (jaune, vert, bleu, mauve). Ils ont ensuite fait la somme de ces distances. Cette mesure rend une donnée pour chacun des points de l'espace associé à un input, donnée qu'ils appellent la « distance interpoint » : sa distance globale aux points causés par les autres inputs. On obtient ainsi un ensemble de paires associant des inputs (rouge, etc.) à la distance interpoint entre le vecteur qu'ils causent en espace d'activation et ceux causés par les autres inputs. Si on regroupe toutes les données des huit réseaux, on obtient une matrice où les colonnes représentent les réseaux (donc huit colonnes, une pour chaque réseau capable d'apprendre la tâche), et où les rangées représentent la distance interpoint pour un input donné (une rangée par input). Si on évalue la corrélation entre les distances interpoints pour un input donné, c'est-à-dire la corrélation des nombres au sein d'une rangée, il s'avère que celle-ci est très élevée, soit au-dessus de 0,9. Malgré leurs différences microscopiques individuelles, les réseaux ont appris à répondre à un input donné en produisant un point dans l'espace d'activation dont la distance avec les autres points est identique (ou similaire, $r > 0,9$). Nous savions par ailleurs que deux réseaux, initialement identiques, entraînés jusqu'à un taux de succès similaire avec le même corpus, mais où les différentes instances du corpus sont présentés dans un ordre différent, produiront des matrices d'interconnexions différentes et, une fois entraînés, répondront par conséquent en produisant des points différents : les deux réseaux possèdent des attracteurs mais ceux-ci sont situés à différents endroits de l'espace. Cette différence reflète les propriétés microscopiques distinctes des réseaux entraînés. Mais nous savons maintenant que ces points seront équidistants des points produits en réponse aux autres inputs. Grâce à l'expérience

36. De la manière dont elle fut construite, l'expérience de Laaksso et Cottrell fait aussi abstraction de la procédure d'apprentissage, de l'environnement d'apprentissage (l'ordre de présentation des données).

de Laaksso et Cottrell, nous savons aussi désormais que ce résultat est indépendant du nombre de dimensions que contient l'espace d'activation de réseau. La très haute corrélation entre les distances interpoints montre que l'apprentissage a amené les réseaux non seulement à créer une partition appropriée de l'espace d'activation mais à créer un « objet » identique (ou similaire) en espace d'activation, c'est-à-dire un « objet » dont les points saillants sont équidistants, et ce, indépendamment du nombre de dimensions de l'espace dans lequel il se trouve³⁷. Il est évidemment difficile d'imaginer visuellement « l'objet » en question. Mais imaginons que nous ayons entraîné un réseau à discriminer seulement deux inputs et que nos dix réseaux aient appris à le faire en associant à chacun un point dans l'espace d'activation. Les résultats de Laaksso et Cottrell suggèrent que les deux points en espace d'activation seront d'une distance égale, ou presque, peu importe le nombre de dimensions de l'espace d'activation du réseau. Il y aura donc une droite de longueur l dans un plan, une droite de longueur l (approximativement) dans un cube, une droite de longueur l (approximativement) dans un hypercube ($n = 4$), une droite de longueur l (approximativement) dans un autre hypercube ($n = 5$), et ainsi de suite. Malgré leurs différences microscopiques importantes (variation du nombre de coefficients synaptiques de deux ordres de magnitudes), les réseaux réussissent la tâche en créant un objet similaire, une droite de longueur l , en espace d'activation. En comparant seulement la distance entre les points, Laaksso et Cottrell font abstraction de la localisation absolue des points dans l'espace de même que des translations, rotations et réflexions possibles des points dans les divers espaces — différences qui sont causées par l'assignation aléatoire initiale des coefficients synaptiques, l'ordre de présentation des stimuli et le nombre de neurones assignés à la tâche. Par cette même mesure, ils font aussi abstraction de la dimensionnalité des espaces dans lesquels ces points se retrouvent, laquelle représente le nombre de neurones au niveau de la couche cachée.

La réalisation de diverses tâches demande la construction d'objets dans l'espace d'activation et ceux-ci sont créés peu importe le nombre de neurones dont dispose le réseau, si on suppose qu'ils en ont suffisamment. Cela suggère qu'à l'intérieur de certains paramètres, le nombre de neurones et le détail de leur interconnexion ne sont pas des variables neurologiques importantes lorsqu'on s'intéresse à la capacité des réseaux de réaliser une même tâche, tout comme la composante directionnelle de la vélocité des molécules de gaz, ou leur vibration et leur rotation, n'est pas pertinente (à l'intérieur de certains paramètres) à la réalisation de la température. Cela suggère qu'il est possible de concevoir des mesures qui font abstraction de ces variables.

37. Rappelons qu'il est toujours possible de placer un objet n -dimensionnel dans un espace contenant un nombre de dimensions égal ou supérieur à n : il est possible de placer un plan ($n = 2$) sur un plan, dans un cube ($n = 3$) ou dans un hypercube ($n > 3$) mais il n'est pas possible de le placer sur une droite ($n = 1$) ou un point ($n = 0$).

Une réponse à l'argument de la réalisation multiple intraspécifique prendra donc la forme suivante. Oui, deux humains peuvent tous les deux croire que le père Noël n'existe pas, même s'ils ne partagent pas le même nombre de neurones et, *a fortiori*, une même matrice d'interconnexion synaptique. Oui, un individu peut continuer à croire que le père Noël n'existe pas malgré la perte quotidienne de neurones et donc la modification quotidienne de sa matrice d'interconnexions. Mais le nombre de neurones et le détail de l'interconnexion ne sont pas des variables neurologiques plus importantes à la réalisation des propriétés psychologiques que la direction des molécules, leur vibration et leur rotation ne le sont à la réalisation de la température. Toute description suffisamment microscopique d'un système sera unique à ce système et impliquera une réalisation distincte d'une propriété macroscopique. Cela montre ou bien qu'il n'existe aucune réduction en sciences, auquel cas l'irréductibilité des propriétés psychologiques n'est qu'une conséquence triviale de l'irréductibilité générale de toutes les propriétés macroscopiques et les scientifiques se leurrent lorsqu'ils croient avoir réduit la température, ou bien que la réduction doit identifier le niveau de description approprié capable de décrire le comportement global des propriétés microscopiques. Tout comme la mécanique statistique permet de décrire le comportement global d'ensembles de molécules d'une manière qui permet la réduction de la température, les neurosciences mathématiques permettent de décrire le comportement global d'ensembles de neurones d'une manière qui permettra la réduction des propriétés psychologiques. Du moins, aucun argument valide ne nous permet désormais de croire le contraire.

3. Conclusion

L'argument inférant l'irréductibilité des propriétés psychologiques à partir de leur réalisation multiple procède d'une attitude foncièrement antinaturaliste à l'égard de la psychologie en lui niant la latitude accordée aux autres sciences. Dans les autres sciences, il est possible de relativiser les réductions à des domaines bien définis, c'est-à-dire des domaines qui découpent la nature d'une manière non *ad hoc*, et de corriger ainsi l'appareil conceptuel des théories utilisées. Dans les autres sciences, il est possible de construire des niveaux abstraits et idéalisés permettant la description du comportement global des systèmes, niveaux qui font abstraction de complexités inutiles. J'ai montré ici que si l'on accorde les mêmes privilèges à la psychologie, la réalisation bel et bien multiple des propriétés psychologiques ne permet pas d'inférer l'irréductibilité du domaine psychique aux propriétés neurologiques ou, plus généralement, physiques des systèmes.

Cela dit, il convient de souligner qu'on commence à peine à comprendre comment construire les niveaux de description appropriés — car il y a en aura certes plus d'un, la psychologie regroupant un amalgame peu unifié de traits humains et animaux — et qu'il revient maintenant aux chercheurs animés du

sentiment réductionniste de construire les niveaux de description appropriés. Les recherches contemporaines, tout comme le contenu du présent article, suggèrent que le niveau de description généralement utilisé pour décrire les autres systèmes complexes, soit la théorie des systèmes dynamiques³⁸, y tiendra une place importante. Comme le remarque Francis Crick, « lorsque nous en arrivons au point où nous pouvons nous asseoir et pondre la dérivation d'une théorie à partir d'une autre, la partie excitante du véritable travail scientifique est terminée³⁹ ». Le travail scientifique excitant en psychologie et neurosciences se poursuit aujourd'hui avec la vigueur renouvelée qu'offre la capacité de modéliser mathématiquement les propriétés microscopiques et macroscopiques des systèmes nerveux. Aucun argument ne laisse désormais croire qu'un philosophe ne pourra pas s'asseoir un jour et pondre la dérivation des diverses théories de la psychologie. Malheureusement pour lui ou elle, la partie excitante du travail sera alors terminée.

38. Voir Keslo, J. A. S., *Dynamic Patterns*, Cambridge (Mass.), MIT Press, 1995 ; Tienson, J. et Horgan, T. E., *Connectionism and the Philosophy of Psychology*, Cambridge (Mass.), MIT Press, 1996.

39. Propos rapportés par Patricia Churchland, *Neurophilosophy*, p. 285. Nous traduisons.

Herméneutique cosmique

Brian McLaughlin

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004956ar>
DOI : <https://doi.org/10.7202/004956ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

McLaughlin, B. (2000). Herméneutique cosmique. *Philosophiques*, 27 (1), 63–76.
<https://doi.org/10.7202/004956ar>

Résumé de l'article

L'herméneutique cosmique est le projet qui consiste à établir tous les faits de notre monde à partir d'un sous-ensemble propre de ces faits — soit des faits de base. Ce texte porte sur un projet spécifique d'herméneutique cosmique qui présuppose que tous les faits de notre monde sont déterminés par ses faits physiques. Suivant Frank Jackson, je formule cette thèse du déterminisme physique de la façon suivante : toute réplique physique minimale de notre monde en est une réplique *simpliciter*. Une réplique physique minimale de notre monde est un monde qui est exactement comme le nôtre sous tous ses aspects physiques, et qui ne contient rien de plus que ce qui est requis pour cela. Une réplique *simpliciter* de notre monde est un monde qui est exactement comme le nôtre sous tous ses aspects. Je me concentre sur la question épistémologique suivante : si la thèse du déterminisme physique est vraie, alors qu'est-ce qu'un démon laplacéen devrait connaître, en plus des faits physiques de notre monde, pour pouvoir prédire tous les faits de notre monde à partir des faits physiques?

Herméneutique cosmique

BRIAN McLAUGHLIN

brianmc@rci.rutgers.edu

Department of Philosophy

Rutgers University

RÉSUMÉ. — L'herméneutique cosmique est le projet qui consiste à établir tous les faits de notre monde à partir d'un sous-ensemble propre de ces faits — soit des faits de base. Ce texte porte sur un projet spécifique d'herméneutique cosmique qui présuppose que tous les faits de notre monde sont déterminés par ses faits physiques. Suivant Frank Jackson, je formule cette thèse du déterminisme physique de la façon suivante : toute réplique physique minimale de notre monde en est une réplique *simpliciter*. Une réplique physique minimale de notre monde est un monde qui est exactement comme le nôtre sous tous ses aspects physiques, et qui ne contient rien de plus que ce qui est requis pour cela. Une réplique *simpliciter* de notre monde est un monde qui est exactement comme le nôtre sous tous ses aspects. Je me concentre sur la question épistémologique suivante : si la thèse du déterminisme physique est vraie, alors qu'est-ce qu'un démon laplacéen devrait connaître, en plus des faits physiques de notre monde, pour pouvoir prédire tous les faits de notre monde à partir des faits physiques?

ABSTRACT. — Cosmic hermeneutics is the project of ascertaining all of the facts of our world from some proper subclass of the facts of our world — some class of ur-facts. I am concerned with a particular project of cosmic hermeneutics that presupposes that all the facts of our world are determined by the physical facts of our world. Following Frank Jackson, I formulate this thesis of physical determinism as follows: any minimal physical duplicate of the actual world is a duplicate *simpliciter* of the actual world. A minimal physical duplicate of our world is a world that is exactly like our world in every physical respect and that contains nothing more than is required to be exactly like our world in every physical respect. A duplicate *simpliciter* of our world is a world that is exactly like our world in every respect whatsoever. I focus on the following epistemological question: if physical determinism is true, then what, in addition to the physical facts of our world, would a Laplacean demon have to know in order to predict all of the facts of our world from the physical facts?

Pierre Laplace a caractérisé, de façon notoire, le déterminisme causal du mouvement relativement à ses conséquences épistémiques pour un observateur suffisamment bien informé qui disposerait d'une capacité de calcul infinie :

Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'Analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux¹.

1. Laplace, 1886, p. vi-vii.

Même si la thèse du déterminisme causal est vraie, pour qu'il connaisse les mouvements passés et futurs de tous les corps, le démon de Laplace devrait non seulement connaître les positions de tous les corps, pour un instant donné, ainsi que toutes les forces dont la nature est animée, mais il devrait aussi savoir quelles sont les propriétés des corps qui génèrent des forces, quelles sont les lois qui régissent les forces, et quelles sont les lois générales du mouvement. Mais même ces connaissances seraient insuffisantes, car le démon devrait également savoir qu'il n'existe aucun autre corps, aucune propriété additionnelle générant des forces, aucune loi additionnelle des forces, et aucune autre loi générale du mouvement. C'est seulement s'il était en possession de ces connaissances additionnelles que le démon serait en mesure de connaître les mouvements passés et futurs de tous les corps.

Terence Horgan a bien saisi la nature du projet dans lequel le démon de Laplace devrait s'engager pour connaître les mouvements de tous les corps, en le caractérisant comme un calcul numérique cosmique². Mais si le démon de Laplace devait aussi connaître tous les faits du monde, alors il devrait s'engager dans un autre projet, soit celui de l'herméneutique cosmique, pour reprendre l'expression de Horgan³. Pour nous en tenir à l'essentiel, et en laissant de côté les subtilités de cette notion, l'herméneutique cosmique consiste à connaître tous les faits de notre monde à partir d'un sous-ensemble propre de ces faits — soit des faits de base si l'on peut dire. Ma discussion dans ce texte porte principalement sur un projet d'herméneutique cosmique encore plus spécifique.

Laplace cherchait à caractériser le déterminisme causal (à tout le moins relativement au mouvement), mais dans ce qui suit, c'est un déterminisme d'un autre ordre qui me préoccupe, à savoir le déterminisme physique, c'est-à-dire la thèse selon laquelle les faits physiques de notre monde déterminent tous les faits du monde. La doctrine du déterminisme physique que je veux discuter est, plus précisément, celle que Frank Jackson appelle « matérialisme », et qu'il formule de la façon suivante :

Toute réplique physique minimale de notre monde est une réplique *simpliciter* de notre monde⁴.

Une réplique *simpliciter* de notre monde est un monde possible qui est exactement comme notre monde sous tous ses aspects. Une réplique physique de notre monde est exactement comme notre monde sous tous ses aspects physiques : c'est un monde qui obéit exactement aux mêmes lois physiques que le nôtre, et qui a exactement le même nombre et les mêmes sortes d'objets et de champs que le nôtre, objets et champs qui sont exactement dans les mêmes états, qui sont impliqués dans les mêmes événements, et qui

2. Voir Horgan, 1983, p. 20, qui utilise l'expression « *cosmic number crunching* ».

3. *Ibid.*, p. 20-21.

4. Jackson, 1998, p. 42.

sont exactement dans les mêmes relations physiques que celles dans lesquelles se trouvent ceux de notre monde.

Mais alors, qu'est-ce qu'une réplique physique minimale de notre monde? Pour être une réplique physique minimale de notre monde, un monde possible doit en être une réplique physique qui ne contient pas plus de lois, d'objets, de champs, de propriétés, d'états, d'événements et de relations qu'il n'est requis pour qu'il soit une réplique physique de notre monde⁵. Bref, pour être une réplique physique minimale de notre monde, un monde doit en être une réplique physique qui ne contient strictement rien de plus que ce qui est requis pour en être une réplique physique.

Un monde peut être une réplique physique de notre monde sans pour autant en être une réplique physique minimale. Pour bien voir cela, il nous suffit d'abord de supposer que notre monde ne contient pas de fantômes, ce qui semble tout à fait inoffensif. Cependant, il pourrait bien y avoir une communauté de fantômes dans un monde qui constitue une réplique physique de notre monde, à condition que ces fantômes n'affectent, causalement, rien de physique dans ce monde. En effet, la totalité des faits physiques de notre monde n'entraîne pas, logiquement, l'inexistence d'une telle communauté de fantômes. Compte tenu de la totalité des faits physiques de notre monde, cette question demeure ouverte, d'un point de vue strictement logique. Puisqu'il n'existe pas une telle communauté de fantômes dans notre monde, comme nous venons de le supposer, et qu'il existe une telle communauté dans le monde possible en question, ce dernier est une réplique physique de notre monde, mais il n'en est pas une réplique physique minimale, puisqu'il contient plus qu'il n'est strictement requis pour être une réplique physique de notre monde, soit la communauté de fantômes en question. Pour prendre un exemple différent, supposons qu'un certain monde possible soit une réplique physique de notre monde, mais qu'en plus d'obéir aux lois physiques de notre monde, il obéisse à des lois additionnelles, qui ne sont pas des lois de notre monde. Cela serait possible à condition que ces lois additionnelles ne modifient en rien le cours des événements physiques de ce monde. Ainsi, alors que ce monde serait une réplique physique du nôtre, il n'en serait pas une réplique physique minimale.

Selon la doctrine du déterminisme physique, toute réplique physique minimale de notre monde est une réplique *simpliciter* de notre monde. En d'autres termes, toute réplique physique de notre monde qui ne contient rien de plus que ce qui est strictement requis pour être une réplique physique de

5. Je suppose que les catégories de lois, d'objets, de champs, de propriétés, d'états, d'événements et de relations (ou un certain sous-ensemble propre de ces catégories) sont les catégories fondamentales. S'il existe des catégories fondamentales additionnelles, nous n'avons qu'à les inclure dans cette liste. L'idée d'une réplique physique minimale est celle d'un monde qui ne contient que des occupants physiques pour chacune des catégories fondamentales d'ingrédients du monde.

notre monde est exactement comme notre monde sous tous ses aspects, quels qu'ils soient. Cette doctrine a pour conséquence que notre monde est une réplique physique minimale de lui-même. Cela peut être clarifié de la façon suivante. Supposons que notre monde ne soit pas une réplique physique minimale de lui-même. De cela, il découle que notre monde contient quelque chose en plus de ce qui est strictement requis pour qu'il soit une réplique physique minimale de lui-même. De sorte qu'un monde pourrait être une réplique physique minimale de notre monde et ne pas en être une réplique *simpliciter*. Le déterminisme physique serait alors faux. Nous voyons donc que si notre monde n'est pas une réplique physique minimale de lui-même, alors le déterminisme physique est faux. Ce qui implique que si le déterminisme physique est vrai, alors notre monde est bien une réplique physique minimale de lui-même.

Évidemment, le déterminisme physique est une thèse empirique : si cette thèse est vraie, elle ne peut être connue qu'*a posteriori*. Que notre monde soit une réplique physique de lui-même est une vérité analytique, mais qu'il soit une réplique physique minimale de lui-même, comme la doctrine du déterminisme physique l'implique, est une hypothèse empirique qui porte à controverse. Le fait que notre monde soit une réplique physique minimale de lui-même, si c'en est un, est un fait logiquement contingent qui n'est pas impliqué, logiquement, par la totalité des faits physiques, ni par la totalité des faits physiques en conjonction avec l'énoncé selon lequel ces faits constituent la totalité des faits physiques de notre monde. Et c'est justement pour cela qu'une réplique physique de notre monde pourrait ne pas en être une réplique physique minimale.

Il est clair que le déterminisme physique est une thèse discutable et sujette à controverse. S'il existe des faits mentaux, sociaux, éthiques et esthétiques, alors si la thèse du déterminisme physique est vraie, tout monde qui est une réplique physique minimale de notre monde devrait également contenir ces faits⁶. Car si cette thèse est vraie, alors toute réplique physique minimale du monde actuel doit contenir exactement les mêmes faits que celui-ci. En termes plus imagés, si le déterminisme physique est vrai, alors, pour créer un monde contenant tous les faits mentaux, sociaux, éthiques, et esthétiques de notre monde, la déesse-mère n'aurait qu'à créer un monde qui est une réplique physique minimale de notre monde. Son travail serait dès lors achevé : elle aurait ainsi créé un monde qui est une réplique *simpliciter* du nôtre et, par conséquent, un monde avec exactement les mêmes faits mentaux, sociaux, éthiques, et esthétiques que ceux de notre monde.

La thèse du déterminisme physique est-elle vraie? Est-il vrai, notamment, que toute réplique physique minimale du monde actuel doit être un monde contenant tous les faits mentaux, sociaux, éthiques, et esthétiques de notre monde? Je suis enclin à répondre positivement à ces questions, et j'ai

6. Voir Jackson, 1998, chap. 1.

plutôt de la sympathie pour cette doctrine. Cela dit, mon but ici n'est pas de tenter de défendre cette doctrine de façon générale, mais c'est plutôt une des conséquences épistémiques du déterminisme physique qui me préoccupe, plus précisément la question suivante : si le déterminisme physique est vrai, alors qu'est-ce qu'un démon laplacéen devrait savoir s'il devait être en mesure de mener à bien le projet de l'herméneutique cosmique?

Supposons que le déterminisme physique soit vrai, et supposons qu'un démon laplacéen connaîtrait tous les faits physiques de notre monde, c'est-à-dire tous les faits physiques nomologiques, et chacun des faits physiques particuliers et généraux de notre monde. La question est alors de savoir ce qu'il devrait connaître en plus de cela pour être en mesure de déduire tous les faits de notre monde.

D'abord, le démon devrait savoir que les faits physiques en question sont les seuls faits physiques de notre monde. Mais il devrait également savoir que notre monde est une réplique physique minimale de lui-même. Sinon, le démon ne saurait pas, par exemple, s'il existe ou non une communauté de fantômes n'ayant aucun effet causal sur les faits physiques. Nous voyons donc que s'il doit être en mesure de prédire tous les faits de notre monde, le démon devrait savoir que notre monde est une réplique physique minimale de lui-même.

De plus, puisque pour n'importe quelle affirmation possible visant à exprimer un fait le démon devrait savoir si ce qu'elle exprime est ou non un fait, le démon devrait posséder la connaissance requise pour comprendre n'importe laquelle de ces affirmations. Cela exigerait, notamment, que le démon possède la connaissance requise pour posséder tout concept nécessaire à la compréhension de chacune de ces affirmations possibles.

De plus, si le démon était lui-même situé dans le monde, il devrait également connaître certains faits indexicaux. Si, par exemple, j'étais moi-même le démon, je devrais savoir que je suis Brian McLaughlin. Car c'est là une connaissance que je ne pourrais acquérir uniquement sur la base de ma connaissance de tous les faits physiques, même si on y ajoutait les connaissances additionnelles que je viens de noter, à savoir que ces faits physiques sont les seuls faits physiques, que notre monde est une réplique physique minimale de lui-même, ainsi que la connaissance de tout ce qui est requis pour comprendre toute affirmation possible sur le monde. En effet, sur la base de ces connaissances il me serait impossible de déduire que je suis Brian McLaughlin. De la même manière, je devrais également savoir que je suis maintenant à tel et tel endroit, à chaque moment où j'existe, tout comme je devrais savoir quel moment est le moment présent, pour chaque moment où j'existe.

Il semble, à première vue, qu'en plus des connaissances que je viens d'énumérer, le démon aurait besoin d'en posséder certaines autres. Par exemple, c'est un fait que l'eau = H_2O . Or, quelqu'un pourrait ignorer ce fait même s'il comprenait l'affirmation que l'eau = H_2O . Car la vérité de cette affirmation ne peut être connue qu'*a posteriori*, de sorte que le démon devrait,

semble-t-il, savoir quelles sortes de phénomènes macroscopiques sont identiques à quelles sortes de phénomènes microscopiques⁷.

Il faut bien voir que l'idée selon laquelle le démon devrait connaître de telles identités théoriques est compatible avec le déterminisme physique. Cette idée serait incompatible avec le déterminisme physique seulement si de tels énoncés d'identité étaient vrais de façon contingente. Cependant, bien que les énoncés d'identité vrais ne puissent être connus qu'*a posteriori*, ils n'en sont pas moins nécessairement vrais car, comme Saul Kripke l'a montré⁸ :

Si $A = B$, alors il est nécessaire que $A = B$ (*Principe de la nécessité de l'identité*).

C'est sur la base de l'indiscernabilité des identiques et de la thèse selon laquelle toute chose est nécessairement identique à elle-même que Kripke a proposé sa célèbre démonstration de ce principe. Il a en outre expliqué clairement en quoi certains énoncés d'identité qui, à première vue, sembleraient être des contre-exemples à ce principe — comme par exemple « Benjamin Franklin est l'inventeur des lunettes à double foyer » — n'en sont pas, dans la mesure où les descriptions sont souvent utilisées de façon non rigide pour désigner différents individus dans différentes circonstances possibles. Comme il l'a démontré, lorsque les termes singuliers de chaque côté du signe d'identité sont des désignateurs rigides, alors si l'énoncé d'identité est vrai, il sera nécessairement vrai. Or, les termes de sortes macroscopiques, comme le terme « eau », par exemple, sont des désignateurs rigides, tout comme le sont les descriptions comme « H_2O ». Que l'eau soit identique à l' H_2O est vrai dans tous les mondes possibles où l'eau existe. Et, de façon plus générale, la connaissance d'identités théoriques n'est pas une connaissance de faits contingents, bien qu'elle soit *a posteriori*. Pour cette raison, la thèse selon laquelle le démon devrait connaître les identités théoriques est compatible avec la doctrine du déterminisme physique.

Considérons maintenant les faits relatifs aux caractères phénoménologiques des expériences, c'est-à-dire les faits concernant les aspects de ce que cela fait que d'avoir certaines expériences, comme par exemple les faits relatifs à ce que cela fait que de faire l'expérience d'une douleur intense. S'il y a de tels faits, alors si le déterminisme physique est vrai le démon laplacéen devrait être capable de les prédire. Appelons cette sorte de faits des faits d'expérience. Si l'une ou l'autre des thèses du behaviorisme analytique ou du fonctionnalisme analytique était vraie des expériences, le démon serait en mesure de prédire les faits d'expérience à partir des connaissances suivantes : i) sa connaissance de la totalité des faits physiques, ii) sa connaissance que cette totalité est la totalité des faits physiques, iii) sa connaissance que notre monde est une réplique physique minimale de lui-même, iv) sa connaissance des faits indexicaux à propos de lui-même (s'il se trouve localisé dans le monde), et finalement v) sa compré-

7. Voir Horgan, 1983, p. 23-24.

8. Voir Kripke, 1980.

hension de toute affirmation visant à exprimer des faits d'expérience. Mais, malheureusement pour le déterministe physique, ces deux thèses sont fausses quant aux faits d'expérience⁹. Par conséquent, les connaissances que nous venons de noter seraient insuffisantes pour que le démon soit en mesure de prédire les faits d'expérience de notre monde.

Si, pour prédire les faits d'expérience de notre monde, le démon devait en outre connaître certaines lois psychophysiques selon lesquelles certains états physiques produisent certains états d'expérience, alors le déterminisme physique serait faux. Mais si, par contre, les caractères phénoménologiques des expériences sont identiques à certaines propriétés physiques ou à certaines propriétés fonctionnelles, alors en connaissant ces identités le démon pourrait prédire les faits d'expérience. La connaissance de ces identités serait *a posteriori*, mais elle n'en serait pas moins une connaissance de vérités nécessaires et, ainsi, le fait qu'une telle connaissance soit requise pour mener à bien le projet de l'herméneutique cosmique serait compatible avec le déterminisme physique.

Convenons d'utiliser l'expression « connaissance des faits physiques de notre monde » en un sens large inhabituel comme une abréviation qui inclut la connaissance des faits physiques de notre monde, la connaissance que ces faits sont les seuls faits physiques de notre monde, et la connaissance que notre monde est une réplique physique minimale de lui-même. Il faut bien noter cependant qu'une telle expression peut être trompeuse, dans la mesure où si le fait que notre monde soit une réplique physique minimale de lui-même peut être appelé un fait physique, il s'agit tout de même d'un fait physique d'une sorte bien particulière. Généralement, les faits physiques sont ceux qui découlent logiquement de la théorie physique qui est vraie de notre monde. Cependant, une telle théorie physique n'entraîne pas que notre monde soit une réplique physique minimale de lui-même, puisque la théorie physique qui est vraie de notre monde est également vraie de toute réplique physique de notre monde ; et, comme nous l'avons noté, un monde peut être une réplique physique de notre monde sans en être une réplique physique minimale. Mais si nous gardons à l'esprit cette distinction importante, il ne devrait plus y avoir de danger de tromper qui que ce soit en utilisant l'expression « connaissance des faits physiques de notre monde » dans le sens élargi que nous venons de stipuler, à savoir comme une abréviation qui inclut la connaissance que notre monde est une réplique physique minimale de lui-même.

David Chalmers et Frank Jackson ont récemment soutenu un certain point de vue qui implique que la connaissance d'identités théoriques portant sur des phénomènes non mentaux — comme par exemple que l'eau = H₂O —

9. Ces deux thèses sont peut-être fausses également de tout type de faits mentaux, mais c'est là une question que je laisse ouverte. Pour des arguments contre le fonctionnalisme analytique, qui me paraissent convaincants (et qui s'appliquent également au behaviorisme analytique), voir notamment Hill, 1991, et Chalmers, 1996.

est une connaissance que le démon pourrait obtenir sur la seule base de la connaissance physique que nous venons de noter, à condition que celle-ci soit complétée par la connaissance requise pour comprendre toute affirmation possible visant à exprimer de telles identités théoriques¹⁰. Jackson et Chalmers reconnaissent qu'on peut comprendre l'affirmation que l'eau = H₂O sans savoir si elle est vraie. Mais, ils maintiennent qu'il serait possible de savoir qu'elle est vraie sur la base de la connaissance requise pour comprendre pleinement toute affirmation possible visant à exprimer des faits, et de la connaissance des faits physiques, dans le sens large que nous venons de stipuler au paragraphe précédent.

Leur raison de soutenir que le démon pourrait connaître des faits comme le fait que l'eau = H₂O sur cette seule base épistémique est la suivante. Même si le concept d'eau n'admet pas une définition *a priori* en termes microphysiques ou en termes neutres (*topic neutral*), il n'en est pas moins associé *a priori* à une description qui en fixe la référence de façon contingente. Une telle description qui fixe la référence correspond approximativement, selon eux, à quelque chose qui ressemble à la description suivante : « le liquide transparent, potable, qui tombe du ciel quand il pleut, qui remplit les océans, les fleuves, les ruisseaux, etc. », même si la façon exacte dont la description devrait être complétée demeure incertaine. Et Chalmers utilise l'expression « la substance aqueuse¹¹ » comme une abréviation d'une telle description qui fixe la référence. Il est important de noter que la description « la substance aqueuse » est une description non rigide (tout comme la description « l'inventeur des lunettes à double foyer »), c'est-à-dire que ce qui satisfait cette description, ce qui compte comme substance aqueuse, peut varier d'un monde possible à un autre. Cependant, on peut facilement rigidifier cette description en y ajoutant tout simplement le mot « actuelle ». Ainsi, la description « la substance aqueuse actuelle » est bien une description rigide, et c'est une vérité nécessaire que l'eau = la substance aqueuse actuelle. L'affirmation que l'eau = la substance aqueuse actuelle correspond à une sorte d'analyse fonctionnelle du concept d'eau. Or, selon Jackson et Chalmers, la connaissance de l'analyse fonctionnelle du concept d'eau serait requise pour avoir une possession complète du concept d'eau, et par conséquent elle serait requise pour une compréhension complète de l'affirmation que l'eau = H₂O. Ils soutiennent que si le démon connaissait tous les faits physiques de notre monde et s'il savait que notre monde est une réplique physique minimale de lui-même, il serait alors en mesure de savoir que l'eau = H₂O, puisqu'il saurait que l'eau = la substance aqueuse actuelle, dans la mesure où une telle connaissance est *a priori*¹². De plus, le démon

10. Voir Chalmers, 1996, et Jackson, 1998.

11. Nous traduisons ainsi l'expression anglaise « *the watery stuff* », utilisée par Chalmers (note de la traduction).

12. Voir Jackson, 1998, chap. 2, et Chalmers, 1996, p. 78-79.

serait en mesure de faire l'inventaire de tous les candidats actuels qui correspondent à la substance aqueuse, et il pourrait savoir qu'il s'agit là de la totalité des candidats actuels. Sur la base de cette connaissance, il pourrait tout simplement déduire que l' H_2O est *la seule* substance aqueuse actuelle. Puisqu'il saurait *a priori* que l'eau = la substance aqueuse actuelle, le démon serait capable de prédire que l'eau = H_2O . Par conséquent, il pourrait mener à bien le projet de l'herméneutique cosmique sans devoir connaître à l'avance les identités théoriques comme l'eau = H_2O . Si Jackson et Chalmers ont raison en ce qui concerne les concepts macroscopiques de sortes naturelles, comme le concept d'eau, alors, si le déterminisme physique est vrai, le démon sera en mesure de déduire les identités macrophysiques/microphysiques à partir de sa connaissance des faits physiques de notre monde et de sa compréhension des affirmations exprimant de telles identités.

Cette idée de Jackson et Chalmers, selon laquelle les concepts de sortes naturelles, comme le concept d'eau, seraient associés *a priori* à des conditions qui en fixent la référence de façon contingente, me paraît relativement plausible. Si de tels concepts ont effectivement des conditions *a priori* contingentes qui en fixent la référence, alors il est tout à fait plausible que la connaissance de ces conditions soit requise pour avoir une compréhension complète de ces concepts. Il me suffit de noter que Jackson et Chalmers appuient de façon assez convaincante l'affirmation voulant que le démon serait en mesure de savoir que l'eau = H_2O uniquement à partir de sa connaissance des faits physiques (au sens large noté plus haut), et de la connaissance requise pour avoir une compréhension complète des affirmations exprimant de telles identités. Quoi qu'il en soit, c'est là une thèse que je n'ai pas l'intention de contester.

Comme Chalmers le souligne, contrairement aux concepts macroscopiques de sortes naturelles, comme le concept d'eau, les concepts phénoménaux — c'est-à-dire les concepts exprimant le caractère phénoménal des expériences — ne sont pas associés *a priori* à des descriptions qui en fixeraient la référence de façon contingente¹³. La condition qui fixe, par exemple, la référence du concept phénoménal de douleur — c'est-à-dire le concept de la qualité phénoménale de douleur — est la qualité phénoménale elle-même telle qu'elle est ressentie, et non pas un mode de présentation contingent de la propriété. Ainsi, les concepts phénoménaux n'admettent pas d'analyses fonctionnelles. Le démon pourrait très bien comprendre, par exemple, que la douleur = l'activation des fibres-C sans être en mesure de déterminer, sur la base de sa connaissance des faits physiques de notre monde, si cette affirmation d'identité est vraie (même si cette connaissance est entendue dans le sens élargi que nous avons noté plus haut). De façon plus générale, selon le point de vue de Chalmers, le démon ne pourrait connaître aucun fait d'expérience uniquement sur la base de sa connaissance des faits physiques de notre monde et

13. Voir Chalmers 1996, chap. 2.

de sa compréhension des affirmations visant à exprimer des faits d'expérience. Cette compréhension et cette connaissance des faits physiques ne lui fourniraient pas une base à partir de laquelle il pourrait déduire un fait d'expérience, quel qu'il soit.

Je suis d'accord avec cette idée de Chalmers selon laquelle les concepts phénoménaux, comme le concept de la qualité ressentie de la douleur, ne sont pas associés *a priori* et de façon contingente à des conditions qui en fixent la référence ; je suis aussi d'accord avec l'idée que la condition qui fixe la référence du concept phénoménal de douleur est la qualité ressentie de la douleur elle-même¹⁴. Et, j'accepte également l'idée que le démon ne pourrait connaître aucun fait d'expérience de notre monde uniquement sur la base de sa connaissance de tous les faits physiques du monde (dans notre sens élargi), et de la compréhension des affirmations exprimant des faits d'expérience.

Cela dit, je suis en désaccord avec le point de vue de Chalmers concernant ce que le démon devrait savoir pour connaître les faits d'expérience de notre monde. La position de Chalmers relativement au projet d'herméneutique cosmique qui nous intéresse ici est la suivante : selon lui, pour mener à bien le projet d'herméneutique cosmique, le démon devrait connaître tous les faits physiques de notre monde, il devrait posséder la connaissance requise pour comprendre toute affirmation possible sur notre monde, il devrait aussi connaître certains faits indexicaux, s'il était localisé dans notre monde, et, finalement, il devrait savoir quels sont les divers faits physiques ou fonctionnels qui sont nomologiquement suffisants pour les divers faits d'expérience. Ainsi, selon ce point de vue, le démon devrait connaître des lois liant les faits physico-fonctionnels aux faits d'expérience, c'est-à-dire des lois phénoméno/physico-fonctionnelles. Si Chalmers a raison de soutenir que le démon devrait connaître de telles lois, alors le déterminisme est faux. Car ces lois ne seraient pas déterminées par les lois physiques et les faits physiques de notre monde. Et, par conséquent, il serait faux de dire que toute réplique physique minimale de notre monde en est une réplique *simpliciter*, puisque les lois phénoméno/physico-fonctionnelles de notre monde seraient différentes de celles de toute réplique physique minimale de notre monde, pour la simple raison que de telles lois feraient défaut à toute réplique physique minimale de notre monde. Chalmers soutient qu'une réplique physique minimale de notre monde serait dépourvue de faits de conscience ; les contreparties de nous-mêmes dans un tel monde seraient des « zombies¹⁵ ». Il soutient donc que le déterminisme physique est faux.

Je pense que Chalmers a tort. Le démon n'a pas besoin de connaître des lois phénoméno/physico-fonctionnelles pour connaître les faits d'expérience

14. Dans la terminologie technique de Chalmers, 1996, chap. 2, l'intension primaire et l'intension secondaire du concept de la qualité ressentie de la douleur coïncident. Je suis d'accord avec ce point.

15. *Ibid.*, chap. 3.

de notre monde. Selon moi, tout ce dont il aurait besoin serait de connaître toutes les affirmations d'identités phéno/physico-fonctionnelles vraies, et il devrait savoir que celles-ci constituent les seules affirmations d'identités phéno/physico-fonctionnelles vraies. Mais, comme je l'ai noté, le fait que ce type de connaissance soit requis pour mener à bien le projet d'herméneutique cosmique est compatible avec le déterminisme physique, car de telles affirmations d'identité utilisent des désignateurs rigides et sont, par conséquent, nécessairement vraies, même si elles sont *a posteriori*. Dans la mesure où elles sont nécessairement vraies, le fait que le démon doit connaître de telles identités est compatible avec le déterminisme physique. De plus, il est très facile de comprendre pourquoi le démon devrait connaître de telles identités. Premièrement, elles ne peuvent être connues qu'*a posteriori*. Deuxièmement, elles ne peuvent être déduites *a priori* à partir de la totalité des faits physiques (dans notre sens élargi), et de la connaissance requise pour comprendre ces faits. Troisièmement, la raison pour laquelle ces identités ne peuvent être ainsi déduites *a priori* est que, contrairement aux macro-concepts ordinaires comme celui d'eau, les concepts phénoménaux n'admettent pas d'analyses fonctionnelles en termes physiques ou en termes neutres (*topic neutral*).

Chalmers a anticipé une objection similaire, et il maintient que quelqu'un qui, pour défendre le déterminisme physique, endosse une position comme celle que je viens de présenter serait engagé à soutenir une thèse qu'il appelle « la nécessité métaphysique forte ». Il caractérise cette dernière de la façon suivante :

Selon cette thèse, il y a moins de mondes métaphysiquement possibles qu'il n'y a de mondes logiquement possibles, et la nécessité *a posteriori* de certaines affirmations peut découler de facteurs qui sont indépendants de la sémantique des termes impliqués¹⁶.

J'ignore si quelqu'un soutient effectivement cette thèse de la nécessité métaphysique forte. Mais, quoi qu'il en soit, ce n'est pas une thèse que j'endosse. Je pense que les mondes métaphysiquement possibles sont tout simplement identiques aux mondes logiquement possibles. Et la nécessité *a posteriori* des identités phéno/physico-fonctionnelles doit être expliquée sur la base de la sémantique des termes qui se trouvent de chaque côté du signe d'identité. Le fait que ces termes soient des désignateurs rigides devra apparaître dans une explication de la raison pour laquelle ces identités sont nécessairement vraies, si elles sont vraies. De plus, le fait qu'un des deux termes exprime un concept phénoménal, alors que l'autre exprime un concept physico-fonctionnel, devra apparaître dans une explication de la raison pour laquelle de telles identités ne peuvent être connues qu'*a posteriori*, bien qu'elles soient nécessaires. Car les concepts phénoménaux ne sont pas définissables en termes de

16. *Ibid.*, p. 137.

concepts physico-fonctionnels ; et, de plus, les concepts phénoménaux n'admettent pas même d'analyses fonctionnelles en termes physico-fonctionnels. C'est d'ailleurs pour cette raison qu'il existe un fossé explicatif, selon l'expression de Joseph Levine, entre le phénoménal et le physico-fonctionnel : les faits et les propriétés d'expérience ne peuvent avoir d'explications réductrices, que ce soit via des définitions ou via des analyses fonctionnelles¹⁷. Les concepts phénoménaux sont spéciaux. C'est d'ailleurs pour cette raison que le problème corps-esprit est un puzzle aussi déroutant depuis Descartes. Cependant, le déterministe physique est en mesure de soutenir que même si les concepts phénoménaux sont spéciaux, les états et les propriétés d'expérience ne le sont pas. Le dualisme conceptuel n'implique pas le dualisme ontologique. Le déterministe physique peut soutenir que de telles affirmations d'identité sont vraies, et nécessairement vraies. Puisque si elles sont vraies, elles le sont nécessairement, le fait que le démon doive connaître de telles identités, pour mener à bien le projet d'herméneutique cosmique, ne pose aucune difficulté à la doctrine du déterminisme physique.

Ce qui sépare un déterministe physique de Chalmers est la question de savoir si les identités phénoméno/physico-fonctionnelles peuvent être vraies. Chalmers semble soutenir qu'elles ne le peuvent pas, et cela sur la base des trois raisons suivantes : (1) si elles sont vraies, alors elles sont nécessairement vraies ; (2) nous pouvons concevoir de façon cohérente qu'elles soient fausses ; et (3) la concevabilité cohérente implique la possibilité, sauf lorsque l'apparence de contingence d'une affirmation d'identité peut être expliquée en montrant que la personne qui prétend concevoir la fausseté d'une telle affirmation aurait confondu le phénomène en question avec des conditions qui fixent, de façon contingente, la référence de l'un ou l'autre des termes. J'accepte (1) et (2), mais je rejette (3). Chalmers ne nous a donné aucune raison convaincante de croire que les identités phénoméno/physico-fonctionnelles ne peuvent être vraies.

Le lien entre la concevabilité et la possibilité est plus faible que celui de conséquence logique. Le fait que *P* soit concevable est une raison *prima facie* de croire qu'il est possible que *P*¹⁸. Mais une telle raison *prima facie* peut être révoquée. Le déterministe physique peut maintenir que la raison *prima facie* de croire, par exemple, que la corrélation entre la douleur et un certain état physico-fonctionnel corrélatif n'est que contingente peut être révoquée sur la base des principes de simplicité théorique et de cohérence générale. Il peut justifier les identités phénoméno/physico-fonctionnelles, dans la mesure où elles offrent la meilleure explication de ces corrélations. De plus, même si j'accepte que nous ne puissions expliquer l'apparence de contingence des affirmations d'identités phénoméno/physico-fonctionnelles de la façon décrite en (3), cela n'est pas la seule façon d'expliquer cette apparence de contingence. Par exem-

17. Voir Levine, 1983, 1998.

18. Voir Yablo, 1993.

ple, dans le cas de l'affirmation que la douleur = l'activation des fibres-C, une telle apparence de contingence s'expliquera du fait que le concept de douleur a un rôle conceptuel très différent de celui du concept d'activation des fibres-C, un rôle conceptuel si différent qu'il ne peut y avoir aucun lien *a priori* entre ces concepts¹⁹.

En conclusion, si le déterminisme physique est vrai, alors, pour mener à bien le projet d'herméneutique cosmique, le démon de Laplace devrait connaître les choses suivantes : la totalité des faits physiques de notre monde, le fait que ces faits physiques constituent la totalité des faits physiques de notre monde, le fait que notre monde est une réplique physique minimale de lui-même, tout ce qui est requis pour comprendre toute affirmation possible sur notre monde, des faits indexicaux sur le démon lui-même, et, finalement, des identités phénoméno/physico-fonctionnelles. Bien sûr, la question de savoir si ces connaissances seraient une base suffisante pour connaître les faits sociaux, éthiques, et esthétiques de notre monde reste ouverte. J'ai de la sympathie pour l'idée qu'elles le seraient, mais la défense de cette thèse devra attendre une autre occasion²⁰.

Bibliographie

- Chalmers, D., 1996. *The Conscious Mind*, Oxford, Oxford University Press.
- Hill, C., 1991. *Sensations: A Defense of Type Materialism*, Cambridge, Cambridge University Press.
- Hill, C., 1997. « Imaginability, Conceivability, Possibility and the Mind-Body Problem », *Philosophical Studies*, 87, p. 61-85.
- Hill, C. et McLaughlin, B. P., 1999. « There Are Fewer Things Than Are Dreamt of in Chalmer's Philosophy », *Philosophy and Phenomenological Research*, 59, p. 445-454.
- Horgan, T., 1983. « Supervenience and Cosmic Hermeneutics », *Southern Journal of Philosophy*, suppl. 22, p. 19-38.
- Jackson, F., 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, New York, Oxford University Press.
- Kripke, S., 1980. *Naming and Necessity*, Cambridge (Mass.), Harvard University Press.
- Laplace, P.-S., 1886. *Théorie analytique des probabilités*, Tome VII des *Oeuvres complètes de Laplace*, 3^e édition publiée sous les auspices de l'Académie des Sciences, par MM. les secrétaires perpétuels, Paris, Gauthier-Villars, 1878-1912.
- Levine, J., 1983. « Materialism and Qualia: The Explanatory Gap », *Pacific Philosophical Quarterly*, 64, p. 354-361.

19. Voir Hill, 1997, Hill et McLaughlin, 1999, et McLaughlin, à paraître, pour un développement et une défense des points proposés dans ce paragraphe.

20. Ce texte a été traduit de l'anglais par Paul Bernier.

Levine, J., 1998. « Conceivability and The Metaphysics of Mind », *Nous*, 32, p. 449-480.

McLaughlin, B. P., (à paraître). « In Defense of New Wave Materialism », dans Loewer, Barry, et Gillett, Carl, dir., *Physicalism and Its Discontents*, Cambridge, Cambridge University Press.

Yablo, S., 1993. « Is Conceivability a Guide to Possibility? », *Philosophy and Phenomenological Research*, 53, p. 1-42.

Qui a peur des *qualia* corporels?

Jérôme Dokic

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004917ar>
DOI : <https://doi.org/10.7202/004917ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Dokic, J. (2000). Qui a peur des *qualia* corporels? *Philosophiques*, 27 (1), 77–98.
<https://doi.org/10.7202/004917ar>

Résumé de l'article

Les *qualia*, considérés comme des propriétés intrinsèques de nos expériences, ne sont pas toujours bien accueillis par les matérialistes, qui préfèrent les considérer comme des propriétés intentionnelles présentées *dans* l'expérience. Je me demande si cette forme de réductionnisme s'applique *aux qualia* de la conscience corporelle. Selon la théorie matérialiste standard, l'objet intentionnel de l'expérience de la douleur, par exemple, est un dommage physique. Mais cette théorie ne rend pas compte de la différence phénoménale entre ressentir une douleur « de l'intérieur » et la percevoir « à l'extérieur ». J'esquisse une autre analyse réductionniste, compatible avec le matérialisme, selon laquelle l'objet intentionnel de la conscience corporelle, contrairement à celui de la conscience externe, dépend de manière constitutive de l'expérience du sujet.

Qui a peur des *qualia* corporels?

JÉRÔME DOKIC

dokic@poly.polytechnique.fr

Université de Rouen et CREA, École polytechnique

RÉSUMÉ. — Les *qualia*, considérés comme des propriétés intrinsèques de nos expériences, ne sont pas toujours bien accueillis par les matérialistes, qui préfèrent les considérer comme des propriétés intentionnelles présentées *dans* l'expérience. Je me demande si cette forme de réductionnisme s'applique aux *qualia* de la conscience corporelle. Selon la théorie matérialiste standard, l'objet intentionnel de l'expérience de la douleur, par exemple, est un dommage physique. Mais cette théorie ne rend pas compte de la différence phénoménale entre ressentir une douleur « de l'intérieur » et la percevoir « à l'extérieur ». J'esquisse une autre analyse réductionniste, compatible avec le matérialisme, selon laquelle l'objet intentionnel de la conscience corporelle, contrairement à celui de la conscience externe, dépend de manière constitutive de l'expérience du sujet.

ABSTRACT. — *Qualia*, conceived as intrinsic properties of experiences, are not always welcome by materialists, who prefer to see them as intentional properties presented *in* our experience. I ask whether this form of reductionism applies to *qualia* of bodily awareness. According to the standard materialist theory, the intentional object of pain experience, for instance, is a bodily damage. This theory, though, is unable to account for the phenomenal difference between feeling pain “from inside” and perceiving it “outside” (seeing oneself or another in pain). I sketch another reductionist analysis which is compatible with materialism, and according to which the intentional object of bodily awareness, unlike that of external perception, constitutively depends on the subject's experience.

1. Les *qualia* et la conscience intentionnelle

« Pour un matérialiste, aucun fait n'est accessible qu'à une seule personne¹ ». Et pourtant, certaines de nos expériences — principalement celles qui relèvent de la conscience perceptive et de la conscience corporelle — semblent avoir une qualité intrinsèque, un « caractère phénoménal » souvent considéré comme étant essentiellement privé. Thomas Nagel (ou plutôt ses traducteurs) parle d'un « effet que cela fait » d'avoir une expérience de ce type². Cet « effet » est déterminé par les *qualia* associés à notre expérience, auxquels il semble que nous ayons immédiatement accès par introspection.

Plutôt que d'éliminer les *qualia*, qui sont profondément enracinés dans notre vie consciente telle que nous la concevons, certains philosophes préconisent de renoncer à la thèse selon laquelle ils sont des propriétés phénoménales intrinsèques de nos expériences³. Fred Dretske écrit à propos de la

1. Dretske, Fred, *Naturalizing the Mind*, Cambridge (Mass.), MIT Press, 1995, p. 65.

2. Nagel, Thomas, « Quel effet cela fait-il d'être une chauve-souris? », dans *Questions mortelles*, traduction P. Engel et C. Engel-Tiercelin, Paris, PUF, 1983.

3. Selon une autre option, l'existence de *qualia* considérés comme des propriétés intrinsèques de l'expérience est compatible avec une vision matérialiste ou naturaliste du monde ;

perception externe que « la qualité de l'expérience, la manière dont les choses nous apparaissent au niveau sensoriel, est constituée par les propriétés de celles-ci telles qu'elles nous sont représentées⁴ ». Ce qui nous apparaît comme des qualités *de* l'expérience sont en réalité des propriétés présentées *dans* l'expérience. Comme ces propriétés sont typiquement extérieures au sujet et à ses états mentaux, le programme qui consiste à réduire les *qualia* à des propriétés ou à des aspects de la scène objective perçue repose sur une forme d'externalisme. Il s'agit en fait d'une forme assez radicale d'externalisme, puisque *toutes* les propriétés de l'expérience immédiatement accessibles à la conscience sont attribuées à la scène objective perçue. Comme le dit Robert Kirk, c'est un programme « ultra-externaliste », puisqu'il a pour but de « donner une explication philosophique de tous les aspects de la conscience uniquement sur la base du contenu des états mentaux déterminé de manière externe⁵ ».

La réussite d'un tel programme est soumise à plusieurs conditions. Elle dépend avant tout de la possibilité d'une théorie matérialiste ou naturaliste de l'intentionnalité⁶. Une autre condition concerne la nature des objets intentionnels eux-mêmes. La réduction des *qualia* à des propriétés présentées dans l'expérience ne sert le matérialisme que dans la mesure où celles-ci sont elles-mêmes objectives et naturelles. Or on a affirmé l'existence de qualités phénoménales représentées comme étant instanciées par les objets eux-mêmes de la scène perçue, et non directement par l'expérience. Ces qualités seraient pourtant constituées, en définitive, par des relations entre ces objets et le sujet conscient, éventuellement par l'entremise de propriétés intrinsèques de son expérience⁷.

voir par exemple Shoemaker, Sydney, *The First-Person Perspective and Other Essays*, Cambridge, Cambridge University Press, 1996, p. 264. Shoemaker rejette également l'idée selon laquelle les *qualia* sont directement accessibles à la conscience ; ils sont connus pour ainsi dire par description. Un doute est alors permis : Shoemaker rend-il vraiment justice à la notion intuitive de *qualia*?

4. Dretske, *Naturalizing the Mind*, p. 1.

5. Kirk, Robert, *Raw Feeling*, Oxford, Clarendon Press, 1994, p. 54. Hormis Dretske, d'autres représentants de ce programme (ou d'une variante plus ou moins radicale de celui-ci) incluent Harman, Gilbert, « The Intrinsic Quality of Experience », dans Tomberlin, James, dir., *Philosophical Perspectives*, vol. 4, Northridge (Calif.), Ridgeview, 1990 ; Tye, Michael, *Ten Problems of Consciousness. A Representational Theory of the Phenomenal Mind*, Cambridge (Mass.), MIT Press, 1995 et Lycan, William, *Consciousness and Experience*, Cambridge (Mass.), MIT Press, 1996.

6. Il y a certes des différences importantes entre ce qui est matériel, ce qui est physique et ce qui est naturel, mais elles ne jouent pas un rôle substantiel dans cet essai.

7. Voir Shoemaker, *The First-Person Perspective*, qui réintroduit de cette manière la notion de *qualia*. Pour une autre conception des qualités phénoménales appliquée à la perception externe, qui ne revendique pas l'existence de propriétés intrinsèques de l'expérience, voir McDowell, John, « Values and Secondary Qualities », dans Honderich, Ted, dir., *Morality and Objectivity*, Boston, Routledge and Kegan Paul, 1985, p. 110-129.

Je laisserai de côté, dans cet essai, le problème de la naturalisation de l'intentionnalité, pour me concentrer sur celui de la réduction des *qualia* corporels à des propriétés intentionnelles (des propriétés présentées dans l'expérience). Par « *qualia* corporels », j'entends les *qualia* associés à la conscience du corps propre, c'est-à-dire à l'expérience directe que nous avons, « de l'intérieur », de notre propre corps et des conditions dans lesquelles il se trouve. Lorsque j'ai mal à la jambe, ou que je ressens un picotement dans le bras, j'ai conscience qu'une partie de mon corps se trouve dans une certaine condition qui m'est ainsi directement accessible. Le cas de la conscience corporelle est particulièrement important pour le défenseur du réductionnisme. Premièrement, cette forme de conscience est incontestablement associée à des *qualia*, à tel point que le langage ordinaire reconnaît plus volontiers l'existence des sensations corporelles que celle des sensations visuelles, auditives ou olfactives. Deuxièmement, la thèse selon laquelle la conscience corporelle a une dimension intentionnelle est plus controversée que dans le cas de la perception externe ; par exemple, certains philosophes considèrent la douleur comme une simple sensation dénuée de signification intentionnelle⁸. Enfin, à supposer que la conscience corporelle soit effectivement une expérience intentionnelle, il reste à montrer que ses objets intentionnels font partie du mobilier du monde matériel.

À mes yeux, la réduction des *qualia* de la perception externe à des propriétés intentionnelles est une stratégie prometteuse. Je tenterai néanmoins de montrer que la conscience corporelle pose des problèmes spécifiques que certains auteurs ont largement sous-estimés. Les conceptions matérialistes actuelles de la conscience corporelle sont soit très incomplètes, soit contraires à l'expérience. Je brosserai à grands traits une analyse de la conscience corporelle qui est en principe compatible avec les aspirations réductionnistes et naturalistes. L'intérêt d'une telle discussion est double. D'une part, elle indique que la conscience corporelle n'est pas un obstacle infranchissable pour qui se réclame du programme matérialiste. D'autre part, elle ouvre la voie à une analyse plus réaliste de la conscience corporelle qui, je l'espère, a un intérêt philosophique intrinsèque.

2. L'objectivité de la conscience corporelle

La conscience corporelle a-t-elle un objet intentionnel? Est-elle une forme de perception objective? Si c'est le cas, la perception consciente est un genre qui comporte (au moins) deux espèces : la conscience externe (la vision, le toucher, l'ouïe, etc.) et la conscience corporelle. Il y aurait donc une essence de la perception plus générale que celle de la perception externe, et qui définit également la conscience corporelle. On peut commencer par dégager trois

8. Voir Searle, John, « Intentionality (1) », dans Guttenplan, Samuel, dir., *A Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994, p. 380.

traits fondamentaux qui participent clairement de l'essence de la perception externe⁹, et qui semblent s'appliquer aussi à la conscience corporelle :

1. Une condition minimale de la perception objective concerne la *localisation spatiale*. Concevoir une expérience objective, c'est avant tout concevoir une expérience spatiale¹⁰. Je peux en principe localiser les objets de la conscience visuelle sur la seule base de mon expérience. Normalement, les objets de la conscience corporelle sont localisés dans le corps propre, ressentis comme ayant une certaine épaisseur spatiale. L'objet de mon expérience corporelle peut être localisé de manière essentiellement vague, comme lorsque je ressens une douleur diffuse dans l'ensemble du corps, mais il est souvent localisé assez précisément, comme lorsque je viens de me piquer le doigt avec une aiguille à coudre. De même qu'il est difficile d'imaginer un son qui ne vienne de nulle part, il est difficile d'imaginer une sensation corporelle qui n'ait aucune signification spatiale.

C'est parce que l'objet de la conscience corporelle se présente comme étant localisé que certains phénomènes caractéristiques de la perception objective sont possibles. Par exemple, je peux avoir une douleur à la main gauche et une autre, *exactement similaire*, à la main droite ; la distinction entre deux douleurs qualitativement similaires mais numériquement distinctes est rendue intelligible par la dimension spatiale de la conscience corporelle. Le rapport constitutif des sensations corporelles à l'espace nous permet également d'envisager une douleur qui se déplace le long du bras ou, pourquoi pas, deux douleurs qualitativement différentes qui se « croisent » sur la surface du corps sans se confondre.

2. Une autre notion fondamentale liée à la perception objective est celle d'*attention*. Il y a sans doute une notion d'attention spécifique à l'expérience de type perceptif, par opposition à la pensée consciente¹¹. Je peux porter mon attention successivement sur différentes régions de mon champ visuel. (Le concept d'attention et celui de champ perceptif sont d'ailleurs étroitement liés.) De même, je peux porter mon attention sur la douleur que je ressens au pied droit, puis sur celle que je ressens à la main gauche. Par ailleurs, porter son attention visuelle sur un objet permet typiquement de recueillir davan-

9. Shoemaker, *The First-Person Perspective*, p. 204-206, propose une liste de huit traits constitutifs de la perception sensorielle (*sense-perception*). Le premier trait mentionné ici ne figure pas explicitement dans sa liste, mais il permet sans doute, avec celui qui concerne l'attention, d'expliquer d'autres traits moins fondamentaux. Par exemple, la perception objective permet de prendre conscience de *faits* parce qu'elle est avant tout une conscience d'*objets* (trait 3 chez Shoemaker) ; elle permet aussi d'avoir des informations qui *identifient* l'objet de la perception (trait 4).

10. Voir Strawson, Peter F., *Les individus*, traduction A. Shalom et P. Drong, Paris, Seuil, 1973, ch. 2, et Evans, Gareth, « Things Without the Mind », dans *Collected Papers*, Oxford, Oxford University Press, 1985.

11. Voir Peacocke, Christopher, « Conscious Attitudes, Attention, and Self-Knowledge », dans Wright, Crispin, Smith, Barry C. et Macdonald, Cynthia, *Knowing Our Own Mind*, Oxford, Oxford University Press, 1998.

tage d'informations sur lui. Il en va de même pour la conscience corporelle. Je peux me concentrer sur ma douleur et lui découvrir, sur la base de mon expérience attentive, de nouvelles propriétés.

3. On pose souvent comme principe que là où il y a perception objective, il doit y avoir la possibilité d'illusions ou d'hallucinations. Si je suis capable de percevoir une table réelle devant moi, je peux également être victime d'une illusion : il n'y a aucune table devant moi (hallucination) ou je perçois la table comme ayant des propriétés qu'elle n'a pas réellement (illusion)¹². Il est vrai que la notion d'illusion ou d'hallucination de la conscience corporelle n'est pas solidement ancrée dans l'usage ordinaire, mais nous sommes au moins enclins à l'utiliser jusqu'à un certain point. Par exemple, le sujet victime d'une douleur « référée » croit ressentir la douleur dans son bras alors qu'elle concerne en réalité son cœur. Le patient qui, après une amputation, a des sensations dans sa jambe fantôme a une hallucination. Enfin, je peux corriger mon jugement à l'égard d'une sensation corporelle — je pensais qu'il s'agissait d'une sensation de pression, et je me rends compte que c'est en fait une douleur légère¹³.

Je supposerai ici que ces traits doivent être pris au sérieux, et qu'il est difficile d'en rendre compte simplement en invoquant la thèse suspecte selon laquelle nous « projetons » dans l'espace des sensations corporelles dénuées de signification spatiale. Toutefois, certains philosophes acceptent la thèse selon laquelle la conscience corporelle est spatiale, mais refusent de la considérer comme une forme de perception objective. Pour la tradition phénoménologique, la conscience corporelle ne présente pas au sujet un état de choses objectif, à la manière de la perception externe. Si la douleur, par exemple, m'est présentée dans un « espace douloureux¹⁴ », elle n'est pas un « objet » dont je pourrais avoir l'expérience à partir de plusieurs points de vue différents, comme lorsque je varie l'angle sous lequel j'observe une table. C'est toujours et nécessairement à partir du même point de vue que je fais l'expérience de ma douleur (celui de mon propre corps), ce qui indiquerait que l'on ne peut pas parler, au sens strict, de perception objective¹⁵.

12. Cette distinction entre illusion et hallucination vient de Austin, John L., *Le langage de la perception*, traduction P. Gochet, Paris, Armand Colin, 1971. Le principe qui lie la capacité de perception véridique à la possibilité d'être victime d'une illusion est compatible avec la « théorie disjonctive de la perception », selon laquelle un acte particulier de perception véridique n'aurait pas pu être illusoire (voir Hinton, James, *Experiences*, Oxford, Clarendon Press, 1973).

13. Voir Armstrong, David, *Bodily Sensations*, London, Routledge and Kegan Paul, 1962, p. 53, et Pitcher, George, « Pain Perception », *Philosophical Review*, 79, 1970, p. 388.

14. Voir Merleau-Ponty, Maurice, *La phénoménologie de la perception*, Paris, Gallimard, 1945.

15. La thèse selon laquelle la conscience corporelle n'est pas une forme de perception objective va de pair avec le refus de considérer le corps propre comme un « objet », qui pourrait entrer et sortir du champ de la conscience comme les objets de la perception externe.

Du point de vue matérialiste, toutefois, la position phénoménologique est une pétition de principe. Si l'on peut montrer que l'objet de la conscience corporelle est de même nature que celui de la perception externe, la même chose pourra en principe être perçue de deux « points de vue » différents. Considérons l'expérience de la douleur. Selon la Théorie Matérialiste Standard (TMS), l'objet intentionnel de cette expérience est un dommage ou un désordre physique relatif à une partie de mon corps plus ou moins délimitée. Lorsque j'ai mal au bras, mon expérience me présente un tort corporel d'un certain type — mon bras est meurtri, blessé, irrité, dérangé. Comme le dit Dretske, « de même que l'expérience visuelle d'un arbre est la conscience (*awareness*) d'un objet non conscient (l'arbre), la douleur est une conscience d'une condition corporelle non consciente (une partie blessée, foulée ou malade)¹⁶ ». Des remarques analogues valent *mutatis mutandis* pour les autres sensations corporelles. La conscience corporelle a un objet intentionnel, mais celui-ci fait partie du domaine d'objets reconnu par le matérialiste.

La TMS peut être motivée par des arguments physiologiques. On peut dire, avec Shoemaker, qu'un trait stéréotypique supplémentaire de la perception objective est le fait que « les croyances perceptives sont causalement produites par les objets ou les états de choses perçus, par l'entremise d'un mécanisme causal qui, normalement, produit des croyances vraies¹⁷ ». Toute perception objective suppose l'existence d'un lien causal et informationnel entre l'expérience et son objet. Sur le plan neurophysiologique, les mécanismes qui sous-tendent l'expérience de la douleur véhiculent de l'information sur certains dommages corporels. Des terminaisons nerveuses, appelées « nocicepteurs », se trouvent dans la plupart des tissus corporels. Les nocicepteurs répondent de manière sélective à des stimuli nocifs — principalement aux déformations mécaniques, à l'augmentation de la chaleur et de la concentration de certaines substances chimiques¹⁸.

Selon la TMS, la conscience corporelle peut être considérée (lorsqu'elle est véridique) comme une forme de perception spatiale d'un état de choses objectif qui concerne le corps propre. Cet état de choses — par exemple, un dommage physique — est en principe accessible à la perception externe. Toutefois, cette théorie se heurte à un obstacle majeur. Il y a sans doute une différence notable (c'est un euphémisme) entre percevoir un désordre physique « de l'intérieur », c'est-à-dire ressentir une douleur, et percevoir le même désordre « à l'extérieur », par le biais de la perception externe. Si les deux for-

16. Dretske, *Naturalizing the Mind*, p. 102. Voir aussi Pitcher, « Pain Perception », Harman, « The Intrinsic Quality of Experience », Tye, *Ten Problems of Consciousness* et Lycan, *Consciousness and Experience*. On trouve des éléments de cette conception chez Reid, Thomas, *Essays on the Intellectual Powers of Man*, edited and abridged by A. D. Woozley, Londres, Macmillan and Co, 1941, p. 165-169.

17. Shoemaker, *The First-Person Perspective*, p. 206.

18. Voir Fields, Howard L. et Price, Donald D., « Pain », dans Guttenplan, dir., *A Companion to the Philosophy of Mind*, p. 452-459.

mes de conscience — corporelle et externe — peuvent partager leurs objets, elles ne présentent manifestement pas ceux-ci *de la même manière*. Le même état corporel objectif est présenté au sujet sous deux modes radicalement différents. De tels modes de présentation ne sont assurément pas bienvenus pour un matérialiste qui préconise la réduction des *qualia* à des propriétés purement intentionnelles. Il y aurait en effet des expériences qui porteraient sur la même scène objective perçue mais qui diffèreraient sur le plan phénoménal. La difficulté pour le matérialiste consiste alors à rendre compte de ces différences sans faire intervenir à nouveau des *qualia* corporels.

La concession qui consiste à accepter l'existence de modes de présentation perceptifs fait ressortir toute la force de la thèse phénoménologique traditionnelle : si la conscience corporelle et la conscience externe présentent la réalité sous des modes radicalement différents, il est naturel de soupçonner qu'en définitive, les objets intentionnels de ces deux formes de conscience ne relèvent pas de la même catégorie ontologique¹⁹. Nous ne sommes pas loin de l'idée selon laquelle il y a un espace réservé aux sensations corporelles, qui n'est pas de même nature que l'espace dans lequel sont perçus les objets de la perception externe²⁰.

L'un des buts de cet article est d'essayer de tracer une voie intermédiaire entre la position de la phénoménologie traditionnelle, assez coûteuse sur le plan ontologique, et la conception matérialiste standard, qui parvient difficilement à expliquer les différences phénoménales entre la conscience corporelle et la conscience externe. Je vais commencer par préciser le sens dans lequel l'objet de la conscience corporelle est présenté d'une manière radicalement différente de celui de la conscience externe.

3. Un modèle discutable de la conscience corporelle

Le problème central de la TMS consiste à rendre compte de la différence intuitive qui existe entre percevoir un dommage physique « de l'intérieur », c'est-à-dire ressentir une douleur, et percevoir le même dommage « à l'extérieur »,

19. On peut ainsi lever l'objection selon laquelle l'existence de modes de présentation perceptifs distincts des objets intentionnels de l'expérience ne représente aucune difficulté pour le matérialiste, en tout cas pas plus que l'existence de termes co-référentiels qui ont des sens (non descriptifs) différents. Certes, si le contenu de l'expérience est entièrement conceptuel (voir McDowell, John, *Mind and World*, Cambridge (Mass.), Harvard University Press, 1994), il n'y a aucune différence de nature entre les deux types de modes de présentation — sur le plan perceptif et sur celui de la pensée conceptuelle. Cette objection ne tient pas dans le cas spécifique de la conscience corporelle. Comme nous le verrons dès la section suivante, les différences entre celle-ci et la conscience externe sont telles qu'il est particulièrement difficile de maintenir la thèse selon laquelle le même état de choses est présenté sous deux modes différents, encore moins si ces modes sont purement conceptuels.

20. Sur ce point, Merleau-Ponty est clair : la distinction entre ces deux genres d'espace est tracée à l'intérieur même de l'expérience. C'est une distinction phénoménologique, alors que la notion d'espace physique est d'ordre scientifique et vient de la réflexion.

par exemple par la vue. Cette différence est liée au fait que l'espace de la douleur ressentie, contrairement à celui des objets vus, a des *bornes* relativement bien définies. Tout d'abord, la douleur telle que je la ressens concerne toujours une partie corporelle, jamais une chaise ou un cendrier. Mais, deuxièmement, elle n'est pas localisée dans n'importe quelle partie corporelle ; elle concerne toujours *mon* corps. Je ne peux pas ressentir une douleur dans un corps qui ne m'appartient pas. Toute douleur que je ressens est localisée dans mon espace corporel.

Le cas du membre fantôme semble constituer à cet égard une exception. Le sujet qui a mal à son bras fantôme ne ressent-il pas une douleur en un lieu extérieur peut-être vide? Même dans ce cas, pourtant, le sujet a l'impression (illusoire) de ressentir la douleur dans une partie du corps qui n'existe plus. La douleur est ressentie comme étant localisée dans son espace corporel *apparent* : le sujet ne se débarrasse pas de sa douleur en quittant la pièce ; la douleur le « suit » dans tous ses déplacements²¹.

Le défenseur de la TMS peut tenter d'affirmer qu'il est contingent, et non pas nécessaire, que la douleur soit toujours ressentie dans l'espace corporel. Il invoquera peut-être Wittgenstein, qui semble avoir sérieusement envisagé la possibilité de ressentir véritablement une douleur dans la main d'autrui, ou même dans un meuble²². Un certain modèle de la conscience corporelle paraîtra alors inévitable. C'est un fait que les nocicepteurs, qui jouent un rôle crucial dans l'expérience de la douleur, se trouvent à l'intérieur ou à la surface du corps, mais il aurait pu en aller autrement. De même que certains procédés artificiels repoussent les limites du champ visuel (une paire de lunettes ou de jumelles), on peut imaginer une prothèse qui prolongerait l'expérience de la douleur au-delà de l'espace corporel. De ce point de vue, la conscience corporelle est *essentiellement* une forme de perception externe ; elle est seulement beaucoup plus « myope », pour ainsi dire, que la vision. Comme la lumière, qui est le milieu de l'information relative à l'objet de la perception visuelle, le système nerveux serait une sorte de milieu de l'information relative au dommage corporel perçu.

Les nerfs eux-mêmes sont remplaçables. On peut imaginer, par exemple, que les nocicepteurs qui se trouvent dans la main de mon voisin sont branchés à un émetteur (en plus d'être rattachés normalement au cerveau de ce dernier). Cet émetteur transmet des informations à un récepteur situé dans mon corps, qui lui-même les fait parvenir à mon cerveau. Si une abeille s'avise

21. Voir Langsam, H., « Why Pains are Mental Objects », *Journal of Philosophy*, 92 (6), 1995, p. 304.

22. Voir Wittgenstein, Ludwig, *Le cahier bleu et le cahier brun*, traduction G. Durand, Paris, Gallimard, 1965, p. 100 et suivantes. Il va sans dire que Wittgenstein n'est pas matérialiste au sens présupposé ici. Il ne considère pas non plus dans ce texte l'expérience de la douleur comme une forme de perception d'un dommage corporel. Il est davantage intéressé par les conditions (fictives) dans lesquelles nous serions amenés à localiser une sensation de douleur dans le corps d'autrui, et parler par exemple du « mal de dent à la dent d'un tel ».

de piquer la main de mon voisin, et si l'information fournie par les nocicepteurs et retransmise à mon cerveau est préservée (autrement dit, si le « bruit » informationnel est réduit au minimum), je ressentirai comme mon voisin une douleur déterminée. Certes, pour rendre ce scénario plausible, il faut invoquer un critère d'identité substantiel du corps propre qui soit indépendant de la conscience corporelle. Nous disposons d'un contre-exemple à l'intuition selon laquelle la douleur ressentie est intrinsèquement liée au corps propre seulement s'il est *indépendamment* établi que la main de mon voisin ne fait pas partie de mon corps (au moment où je ressens la douleur en question).

L'intention du matérialiste est ici de montrer que les *qualia* corporels apparemment responsables de l'effet que cela fait de ressentir la douleur « de l'intérieur » ne sont pas des propriétés essentielles de l'expérience corporelle. Cela revient à *éliminer* les *qualia* de ce genre s'il prétend en plus que le contraste intuitif qui existe entre la conscience externe et la conscience corporelle est illusoire.

La situation est en réalité plus complexe, puisqu'il faut distinguer deux cas de figure. Dans le premier cas, lorsque je ressens une douleur dans la main de mon voisin, je n'ai pas conscience que cette main fasse partie de mon corps propre ; j'ai même conscience qu'elle n'en fait pas partie. Autrement dit, il y aurait une dissociation entre la conscience des sensations corporelles éventuellement localisées dans l'espace extracorporel et la conscience « de l'intérieur » des limites spatiales de mon corps propre, qui serait indépendante des sensations corporelles. Dans le deuxième cas de figure, lorsqu'il me semble ressentir la douleur dans la main de mon voisin, j'ai également conscience que celle-ci fait partie de mon corps propre : j'ai l'impression d'avoir une troisième main. Pourtant, ma conscience « de l'intérieur » des limites de mon corps propre, contrairement à mon expérience de la douleur, est illusoire ; *en réalité*, la douleur que je ressens concerne un corps qui ne m'appartient pas. Dans ce cas aussi, il y a une dissociation entre la conscience des sensations corporelles et celle du corps propre, puisque les deux types d'expérience n'ont pas les mêmes conditions de satisfaction. Toutefois, à la différence du premier cas, cette dissociation n'est pas apparente sur le plan phénoménologique. Aucun de ces cas de figure n'est favorable au matérialiste, puisqu'ils reviennent tous deux à créditer la conscience « de l'intérieur » des limites du corps propre des *qualia* qui ont été refusés à la conscience corporelle.

En résumé, il n'est pas aisé d'éliminer les *qualia* corporels responsables de l'effet que cela fait de ressentir une sensation « de l'intérieur ». Ces *qualia* semblent être essentiels à la conscience corporelle, et participer de l'impression que nous avons des limites de notre corps propre.

4. Un critère d'identité partiel du corps propre

Le matérialiste n'est pas obligé de prendre au sérieux le modèle de la conscience corporelle esquissé dans la section précédente ; il peut tenter de rendre

justice à l'intuition selon laquelle la douleur ressentie est intrinsèquement liée à l'espace corporel. Les limites de la conscience corporelle sont *par définition* celles du corps propre. Il est certes possible que nos nocicepteurs se trouvent en dehors de ce qui est *de facto* notre corps. Mais lorsque nous envisageons cette possibilité, nous ne considérons pas que la douleur est ressentie dans l'espace extra-corporel ; nous supposons au contraire que les limites de cet espace se sont déplacées, et que le point auquel nous ressentons la douleur (dans la situation envisagée) fait maintenant partie intégrante de *notre* corps. Comme le dit Armstrong, « c'est un fait contingent que les qualités sensibles [telles que la douleur] concernent les objets qu'elles concernent et pas d'autres, mais ce n'est pas un fait contingent que ces objets sont appelés les *corps* des personnes²³ ».

Selon la définition envisagée, une condition suffisante pour qu'un membre donné appartienne au corps propre d'un sujet est que celui-ci ressent ou pourrait ressentir une sensation corporelle localisée dans ce membre. Ce ne peut pas être une condition nécessaire de la possession du corps propre, du moins si nous admettons que nous avons beaucoup de parties corporelles dont nous ne sommes normalement pas conscients. En continuant à prendre la douleur comme paradigme de la sensation corporelle, nous obtenons le critère d'identité partiel suivant :

(D) Si j'ai mal (ou si je pouvais avoir mal) à x , alors x fait partie de mon corps propre, ou inclut une limite de celui-ci.

Une difficulté immédiate liée à ce critère concerne les conditions de satisfaction de la conscience corporelle. Il ne s'agit pas d'affirmer que si je ressens *apparemment* une douleur en un endroit déterminé de l'espace, celui-ci est occupé par une partie de mon corps propre. Une telle affirmation conduirait à la définition d'un corps purement phénoménal (éventuellement constitué en partie de membres fantômes), qui ne pourrait être identifié à aucun objet matériel. De plus, l'objet intentionnel de l'expérience de la douleur ne pourrait plus être identifié dans tous les cas à un dommage physique. Selon une lecture de (D) plus acceptable pour le matérialiste, si je ressens *véritablement* une douleur en un endroit déterminé de l'espace, celui-ci est occupé par une partie de mon corps propre. Le problème consiste maintenant à définir ce qui compte comme une expérience corporelle véridique. La simple existence d'un lien causal/informationnel entre un dommage physique et une expérience corporelle ne suffit pas à rendre celle-ci véridique. En l'absence d'un critère de véridicité de la conscience corporelle, le critère (D) apparaîtra entièrement *ad hoc* au philosophe qui, comme Wittgenstein, admet la possibilité de ressentir véritablement la douleur dans un corps étranger. Comment distinguer le cas possible où je localise de manière illusoire la douleur dans un corps étranger de celui où mon corps propre incorpore (temporairement) la main d'autrui?

23. Armstrong, *Bodily Sensations*, p. 48.

Il est intéressant de comparer le critère (D) à ce que Quassim Cassam appelle « l'idéalisme concernant la possession du corps²⁴ ». Selon cette position, attribuée à Locke, notre corps est entièrement défini comme celui dont nous avons conscience de la manière appropriée, c'est-à-dire « de l'intérieur ». Une partie corporelle appartient au corps du sujet *si et seulement si* le sujet ressent ou peut ressentir des sensations corporelles dans cette partie. Les sensations corporelles constituent le *seul* critère adéquat d'identité du corps propre.

Le critère (D) n'est pas idéaliste au sens de Cassam, avant tout parce qu'il ne fait pas de la conscience corporelle une condition nécessaire de l'appartenance d'un membre donné au corps propre du sujet. Mais considérons l'objection que Cassam soulève contre l'idée selon laquelle la conscience corporelle est une condition suffisante de la possession du corps. Revenons au scénario dans lequel la main de mon voisin est branchée à mon cerveau par l'entremise d'ondes radio. Supposons que je ressente une douleur que je localise, sur la seule base de mon expérience, dans cette main. Selon l'idéalisme concernant la possession du corps, il s'ensuit que la main appartient à mon corps propre. Selon Cassam, une telle conclusion est inacceptable car « elle ne respecte pas l'intuition selon laquelle un membre auquel on n'est pas matériellement uni ne peut pas faire partie de soi²⁵ ». Le principe invoqué est ici que « l'on ne peut pas être matériellement uni à un membre spatialement éloigné, auquel on n'est pas localement rattaché²⁶ ».

Pour rendre justice à cette intuition, Cassam oppose à l'idéalisme une position qu'il appelle « réaliste », selon laquelle « il est nécessaire et suffisant, pour qu'un membre ou une partie corporelle fasse partie de nous, que ce membre ou cette partie soit matériellement uni à nous²⁷ ». Bien entendu, le réalisme concernant la possession du corps soulève immédiatement la question de savoir ce qui fait qu'un corps donné, matériellement uni, constitue *notre* corps. Mais je laisserai ce point de côté. Ce qui m'intéresse, c'est la possibilité pour le matérialiste de restreindre les conditions de véridicité de l'expérience de la douleur à celles où l'objet intentionnel est un dommage physique *relatif à un corps matériellement uni*. Le défenseur de la TMS pourrait donc affirmer, en invoquant le réalisme au sens de Cassam, qu'une expérience corporelle est illusoire si le dommage physique avec lequel elle a un lien causal/informationnel concerne une partie matériellement distincte du reste du corps. Cette nouvelle définition exclut du domaine des expériences corporelles véridiques la conscience du sujet dans le scénario wittgensteinien.

La stratégie matérialiste dépend ici du bien-fondé de la position que Cassam qualifie de « réaliste ». Le réalisme exclut la possibilité que les parties de notre corps propre soient éparpillées dans le monde. Or il n'est pas du tout

24. Cassam, Quassim, *Self and World*, Oxford, Clarendon Press, 1997.

25. *Ibid.*, p. 65.

26. *Ibid.*

27. *Ibid.*, p. 66.

évident que cette possibilité soit toujours problématique, comme semble l'affirmer Cassam. Même si cette possibilité n'est en fait jamais réalisée, nous pouvons nous imaginer sans incohérence qu'elle le soit. Supposons par exemple que nous puissions volontairement détacher notre main de notre bras pour explorer tactilement une région de l'espace qui nous est autrement inaccessible. (Il faut imaginer que la main possède son propre mécanisme de locomotion.) Sans me lever ni déranger mes compagnons, je pourrais « détacher » ma main et aller chercher la salière se trouvant à l'autre bout de la table²⁸.

Le scénario qui vient d'être évoqué est curieux, mais pas contradictoire. De manière générale, il n'y a aucune absurdité conceptuelle dans l'idée que le sujet peut agir, peut-être par télékinésie, en des points disjoints de l'espace. À cet égard, il est intéressant de relever que la psychologie expérimentale reconnaît l'existence de modules perceptivo-moteurs spécialisés. Les traitements spatiaux requis pour l'action sont *fragmentés* en multiples systèmes de contrôle relativement indépendants entre eux. Par exemple, différentes informations perceptives et différents groupes moteurs sont impliqués lorsque le bras se déplace vers un objet, et lorsque les doigts le saisissent²⁹.

Il est vrai que notre conscience spatiale, corporelle et externe, est normalement *unifiée*. L'unité de l'expérience découle directement de l'exigence générale selon laquelle le sujet doit pouvoir localiser l'objet perçu sur la seule base de son expérience. Pour que le sujet puisse percevoir un objet comme étant à un endroit déterminé de l'espace, il doit pouvoir le localiser relativement à tous les autres objets qu'il perçoit au même moment, y compris le corps et ses différentes parties. Le champ perceptif est un système de relations spatiales *synesthésiques*, c'est-à-dire entre objets qui peuvent être perçus au travers de différentes modalités sensorielles³⁰.

Selon le matérialisme, la conscience corporelle est une forme de perception objective. Mais pour que ma main détachée puisse contribuer à mon expérience globale du monde, les données qu'elle fait apparaître doivent être *orientées* par rapport au reste des données sensorielles dont j'ai l'expérience. L'orientation relative est une condition nécessaire pour que je puisse placer les informations véhiculées par mon expérience dans un espace dans lequel je me trouve également, et donc pour que je puisse me servir de ma main pour percevoir le monde qui m'entoure.

28. Sur la main détachable, voir aussi Casati, Roberto et Dokic, Jérôme, *La philosophie du son*, Nîmes, Éditions Chambon, 1994.

29. Voir Rossetti, Yves, « Des modalités sensorielles aux représentations spatiales en action : représentations multiples d'un espace unique », dans Proust, Joëlle, dir., *Perception et intermodalité. Approches actuelles de la question de Molyneux*, Paris, PUF, 1997. Je ne veux pas exclure qu'il y ait des actions qui engagent l'ensemble du corps, comme le déplacement d'un endroit à l'autre, ni que ces actions jouent un rôle important dans la constitution de l'objectivité de notre expérience.

30. Voir Dokic, Jérôme, « La signification des expressions égocentriques », dans Proust, dir., *Perception et intermodalité*.

En d'autres termes, ma main reste intégrée au reste de mon corps seulement si elle est spatialement reliée à celui-ci dans mon champ perceptif global. Tant que je garde la trace perceptive de l'orientation de la main par rapport au reste de mon corps, j'ai une expérience corporelle bizarre peut-être, mais unifiée. Je peux considérer ma main comme une partie spatialement disjointe de mon corps.

Que se passe-t-il si je *perds* la trace de l'orientation relative de la main? Ce serait le cas si, par exemple, ma main détachée disparaissait de mon champ visuel pendant un certain temps — si je quittais la pièce en oubliant de rattacher ma main au reste de mon corps. Ce cas serait similaire, sous certains aspects, à la situation dans laquelle se trouvent certains sujets qui ont perdu presque entièrement le toucher et la proprioception musculaire³¹. Ces patients peuvent ressentir la douleur et avoir certaines sensations thermiques, mais ils sont incapables de pointer du doigt l'endroit stimulé sur leur corps s'ils ne le *voient* pas également. Par contre, ils peuvent nommer l'endroit en question et le désigner sur une représentation schématique du corps, ce qui indique que leur image corporelle n'a pas entièrement disparu.

On peut observer que ces patients ne localisent pas le dommage corporel *sur la seule base de leur expérience de la douleur*. Cette expérience, contrairement à la nôtre, ne leur fait pas apparaître les relations spatiales synesthésiques qui existent entre la partie corporelle douloureuse et les parties corporelles visibles. Ils sont contraints de passer par une représentation anatomique de leur corps pour retrouver l'endroit douloureux dans leur champ de vision. De même, il est possible que je sois capable de nommer la partie douloureuse de ma main — que je sache que c'est l'index qui me fait mal, par exemple. Mais cela n'implique pas que mon expérience de la douleur soit véridique. Elle ne pourra pas l'être si je suis incapable de localiser l'index dans mon champ perceptif sur la seule base de mon expérience.

Il est possible aussi que je garde l'impression que ma main est orientée d'une certaine façon relativement au reste de mon corps. Ainsi, j'ai l'impression de percevoir certaines données tactiles à un endroit déterminé dans mon champ perceptif global. Mais cette impression ne peut pas être véridique si le lien nomologique entre l'orientation apparente et l'orientation véritable est rompu. Même si l'orientation apparente de la main correspondait à son orientation véritable, ce serait un pur accident et la main ne serait pas véritablement perçue.

Une possibilité beaucoup plus spéculative, mais qui ne peut pas être écartée *a priori*, est que la main détachée devienne un *autre* corps propre, littéralement aveugle. J'aurais ainsi la capacité de « sauter » d'un corps dans

31. Cole, Jonathan et Paillard, Jacques, « Living without Touch and Peripheral Information about Body Position and Movement : Studies with Deafferented Subjects », dans Bermúdez, José Luis, Marcel, Anthony et Eilan, Naomi, dir., *The Body and the Self*, Cambridge (Mass.), MIT Press, 1995.

l'autre, et de me retrouver instantanément en des lieux complètement différents. Il n'est pas nécessaire d'explorer cette possibilité plus avant. Il suffit de souligner que les considérations qui précèdent ne portent pas contre la possibilité d'avoir différents corps à différents moments, mais plutôt contre la possibilité de jouir de deux points de vue *simultanés* sur le monde qui ne soient pas spatialement reliés l'un à l'autre dans un champ perceptif global. Le don d'ubiquité en ce sens n'existe pas³².

En résumé, une condition relative à l'unité de l'expérience doit effectivement être incluse dans les conditions de véridicité de la conscience corporelle (et donc dans la définition du corps propre), mais elle découle d'une condition plus générale qui concerne l'unité spatiale et synesthésique du champ perceptif. Corrélativement, elle n'implique pas directement l'unité matérielle du corps. On objectera que pour que les parties de mon corps puissent transmettre des informations à mon cerveau, il faut qu'elles soient liées entre elles au moins par une relation causale. Cette objection est sans doute correcte, mais elle montre seulement que nous concevons difficilement la possibilité d'une communication immédiate entre deux points disjoints dans l'espace, qui ne dépende pas d'une transmission causale de l'information d'un point à l'autre continûment. La difficulté conceptuelle que nous avons ici n'est pas spécifique au problème du corps propre.

5. Un bref rappel des options

Dans les deux dernières sections, j'ai soulevé un certain nombre de difficultés liées à la TMS, qui ont trait au contraste intuitif entre la conscience externe et la conscience corporelle. On a distingué deux options. Selon la première, les sensations corporelles ne sont pas nécessairement localisées dans le corps propre. Comme nous l'avons vu, cette option ne permet pas d'éliminer les *qualia* corporels, qui réapparaissent associés à une forme de conscience du corps propre détachée des sensations corporelles (et dont l'existence est douteuse). La deuxième option, qui a l'avantage de respecter l'intuition ordinaire, consiste à définir le corps propre au moyen des sensations corporelles. Non seulement cette option est, nous l'avons vu, *ad hoc*, mais elle ne fait que mettre en évidence le problème qui se pose pour la TMS. Si l'objet intentionnel de la conscience douloureuse n'est qu'un simple dommage physique, il est légitime de se demander ce qu'elle a de spécial par rapport à la conscience externe du même dommage, en ce qu'elle définit les limites du corps propre. Le critère (D) ne fait que souligner le fossé qui existe entre ces deux formes de conscience. Si avoir mal, c'est percevoir un dommage corporel qui, par essence, n'est pas lié à un corps particulier, un tel critère apparaît comme étant tout à fait immotivé.

32. Voir Strawson, Peter F., *The Bounds of Sense*, Londres, Routledge and Kegan Paul, 1966, p. 150-152, et Evans, Gareth, *The Varieties of Reference*, Oxford, Blackwell, 1982, ch. 6.

Pour sortir de cette impasse, je suggère de renoncer à l'idée selon laquelle l'objet intentionnel de l'expérience de la douleur est un simple dommage physique. Dans le reste de cet essai, je vais procéder en trois étapes. Premièrement, je vais rappeler un argument selon lequel les sensations corporelles engagent de manière constitutive des réponses spontanées de la part du sujet. Deuxièmement, je vais montrer que ces réponses participent de l'objet intentionnel de la conscience corporelle. Enfin, je vais tenter de mettre en évidence la dimension réflexive de la conscience corporelle, qui selon moi contribue à expliquer la différence pertinente entre ressentir la douleur « de l'intérieur » et la percevoir « à l'extérieur ».

6. Les conditions d'émergence de la conscience objective

La thèse selon laquelle la conscience corporelle engage nécessairement des réponses spontanées est la conséquence d'un argument plus général présenté par David Pears (à partir de Wittgenstein) et avant lui par Friedrich Hayek³³. L'argument de Hayek/Pears ne concerne pas la conscience corporelle en particulier, mais toute forme d'expérience intentionnelle de type sensoriel³⁴.

En discutant les objections de Wittgenstein contre la possibilité d'un langage privé, Pears affirme que « le langage doit se surimposer à un réseau préexistant de connexions allant des stimuli physiques jusqu'au seuil de la conscience, et en retour jusqu'au comportement³⁵ ». Lorsque l'enfant introduit dans son langage un nouveau terme qui désigne une sensation, il le fait toujours sur un arrière-plan de liens prélinguistiques complexes entre des stimulations sensorielles et des réponses motrices. Ces liens relèvent de ce que Pears appelle des « lignes de types ». Ces lignes de types sont solidement établies avant l'avènement du langage, et prédéterminent le sens et l'usage du terme de sensation.

L'existence de connexions prélinguistiques entre des stimuli et des réponses est particulièrement évidente dans le cas de la douleur. Chez une personne normale, les stimuli nocifs sont associés naturellement à un comportement de rejet approprié à la situation — retirer brusquement sa main d'un radiateur brûlant en est un exemple particulièrement visible. L'idée générale de Pears est que le *type* de douleur ressentie est déterminé par une ligne qui va de certains stimuli à certaines réponses spontanées. Si je veux introduire un terme qui désigne un certain type de douleur, je dois tenir compte des deux moitiés de la sensation,

33. Quelques paragraphes de cette section et de la suivante sont tirés de mon article « La perception interne et la critique du langage privé », *Revue de théologie et de philosophie*, 130, 1998, p. 1-19. Je parviens ici à une conclusion similaire, mais par une voie un peu différente.

34. Un argument similaire, qui concerne la localisation dans le champ perceptif, a été présenté par Pitcher, George, *A Theory of Perception*, Princeton, Princeton University Press, 1971 et Evans, *The Varieties of Reference*.

35. Pears, David, *La pensée-Wittgenstein : du Tractatus aux Recherches Philosophiques*, traduction C. Chauviré, Paris, Aubier, 1993, p. 371.

réceptrice et motrice. Je ne peux pas introduire un terme de sensation privée qui identifierait la douleur indépendamment de son expression naturelle³⁶.

Suivant Pears, les qualités accessibles au sujet d'une expérience consciente sont différenciées les unes des autres par des réponses naturellement associées à certaines stimulations sensorielles. L'idée est que des éléments perçus qui diffèrent sur le plan phénoménal doivent correspondre à des réponses publiquement observables différentes. Dans le cas précis, ces réponses déterminent non seulement le lieu, l'étendue et l'intensité de la douleur ressentie, mais préfigurent également son caractère intrinsèquement désagréable. Les réponses spontanées sont déjà des amorces de rejet de la partie corporelle stimulée, ou du moins des tentatives d'échapper à l'agression extérieure³⁷.

Hayek présente un argument similaire à celui de Pears, sauf qu'il concerne non pas l'introduction d'un nouveau terme de sensation, mais l'émergence d'une expérience objective consciente. Selon lui, les stimuli sensoriels sont associés à des réactions spontanées, et cette association naturelle est antérieure à toute expérience objective. Bien que Hayek ne mentionne pas explicitement l'expérience de la douleur, ses remarques sont applicables à ce cas spécifique. Sur le plan physiologique, lorsque l'information véhiculée par les nocicepteurs parvient au système nerveux central, elle est déjà investie d'une certaine « signification » pour l'organisme. Le système nerveux central reçoit presque simultanément deux « rapports » concomitants : l'un concernant le stimulus nocif, l'autre ayant trait à la réaction spontanée du corps³⁸.

Les lignes de types sont solidement établies avant l'avènement du langage (Pears) et l'émergence d'une expérience consciente (Hayek), mais des connexions régulières entre les stimuli nocifs et les réponses spontanées sont nécessaires pour maintenir à travers le temps la signification du terme de sensation et le contenu intentionnel de l'expérience³⁹. Toutefois, les lignes de types acquièrent un statut différent lorsque le langage et l'expérience sont en place.

36. Voir Wittgenstein, Ludwig, *Investigations philosophiques*, traduction P. Klossowski, Paris, Gallimard, 1961, §245 : « Comment puis-je aller jusqu'à vouloir intervenir au moyen du langage entre la douleur et son expression? »

37. Je suppose ici que l'expérience de la douleur est intrinsèquement désagréable indépendamment des attitudes émotionnelles qui lui sont typiquement associées. On peut contester cette supposition : des patients lobotomisés ou ayant pris de la morphine disent ressentir la même douleur avec la même intensité qu'avant l'opération ou la prise de la drogue, mais affirment qu'elle ne les gêne plus ; leur expérience de douleur n'est plus désagréable, ce qui implique peut-être l'absence de réactions négatives spontanées (voir Dennett, Daniel, « Why You Can't Make a Computer that Feels Pain », dans *Brainstorms*, Cambridge (Mass.), MIT Press, 1978). On peut débattre de la question de savoir s'il faut prendre ou non les dires de ces patients à la lettre. Cette question est relativement indépendante de mon propos dans cet essai. Même si l'expérience de douleur n'est pas intrinsèquement désagréable en ce sens, elle reste liée à des réponses spontanées qui déterminent le type de sensation ressentie — le lieu, l'étendue et l'intensité de la douleur.

38. Hayek, Friedrich, *The Sensory Order*, Chicago, University of Chicago Press, 1953, p. 79-101.

39. Voir Pears, *La pensée-Wittgenstein*, p. 407.

Au commencement, on ne trouve que des stimulations sensorielles et des réponses motrices dénuées de signification intentionnelle. Mais les deux termes des lignes de types prennent une autre dimension avec l'avènement du langage et l'émergence de la conscience objective. D'une part, le lien régulier entre un dommage corporel et le stimulus nocif est tel que l'expérience de douleur peut être considérée comme une forme de perception intentionnelle (confuse) du dommage. D'autre part, les réactions aux stimuli nocifs, bien qu'elles restent spontanées, acquièrent également un statut proto-intentionnel. Premièrement, elles peuvent être interprétées comme des réponses aux dommages eux-mêmes comme stimuli « distaux » (bien que normalement à l'intérieur du corps), et non pas à des stimuli nocifs « proximaux » (autour des nocicepteurs). Deuxièmement, chez l'individu développé, les connexions prélinguistiques entre stimuli et réponses sont suivies et orientées par des comportements supérieurs, plus directement contrôlés par lui. Ces comportements sont capables, dans une certaine mesure, d'inhiber ou de modifier les réponses directes dans le contexte plus général d'autres perceptions et intentions⁴⁰. Le caractère proto-intentionnel des réponses est défini précisément par le fait qu'elles sont sensibles à un tel contexte. Même si en un certain sens, elles ne sont pas entièrement voulues par le sujet, elles représentent quelque chose que nous faisons naturellement, et pas seulement quelque chose qui est fait en nous. Même les mouvements réflexes sont typiquement influencés par la perception de la situation générale : je peux lâcher une tasse de café brûlante dans un contexte et la reposer maladroitement sur une table proche dans un autre⁴¹.

Lorsqu'il n'y a aucune réponse observable, comme dans le cas d'une douleur faible, les stimuli nocifs sont quand même associés par nature à une certaine *tendance* à réagir « visiblement » (par exemple par une estimation verbale de l'intensité de la douleur — mais d'autres manifestations sont possibles). Toutefois, cette tendance n'est pas une simple disposition ; elle constitue déjà un début d'activité réelle qui imprime une direction spécifique à tout comportement fondé sur l'expérience de la douleur. Dans tous les cas, il doit y avoir une certaine *harmonie* entre les réponses spontanées et les comportements motivés par des attitudes à l'égard de la douleur, telles que le désir qu'elle cesse immédiatement⁴². Précisément en raison du fait que ces attitudes sont fondées sur l'expérience consciente de la douleur plutôt qu'antérieures à celle-ci, elles en sont dissociables dans un contexte plus large où d'autres considérations interviennent — par exemple, une disposition au masochisme. Même le masochiste, pourtant, ne peut pas aller entièrement contre sa nature, et doit résister aux réponses spontanées s'il ne les souhaite pas.

40. Voir Hayek, *The Sensory Order*, p. 85.

41. Voir Merleau-Ponty, Maurice, *La structure du comportement*, Paris, PUF, 1942 ; Melzack, R. et Wall, P., *The Challenge of Pain*, London, Penguin Books, 1988 (deuxième édition), p. 193.

42. Voir Evans, *The Varieties of Reference*, p. 159.

7. Une définition des sensations corporelles

Selon l'argument wittgensteinien de Pears, les règles sémantiques qui gouvernent l'usage de *tous* les termes de sensations — que celles-ci relèvent de la conscience externe ou de la conscience corporelle — sont édifiées sur un sol naturel de liens préreflexifs entre la perception et le comportement. Il existe néanmoins une différence cruciale entre les deux formes de conscience, que l'on peut présenter de la manière suivante. Il se trouve qu'un terme comme « rouge » désigne une propriété qui convient naturellement à la cause « distale » de notre expérience chromatique. Sa *référence* est (un aspect de) l'*entrée* de la ligne de type qui sous-tend son utilisation, même si son *sens* dépend de la mise en place préalable de la ligne tout entière. Comme il n'y a pas de comportement caractéristique associé à la perception d'un objet rouge, la relation entre les stimuli et les réponses qui sous-tendent la sensation de rouge est complexe et indirecte, et doit être conçue sur un modèle holiste⁴³. Par contraste, les liens qui correspondent à une sensation comme la douleur sont si simples et directs que le terme « douleur », compris dans son sens ordinaire, ne peut pas être utilisé *pour désigner la cause de la douleur indépendamment de son expression naturelle*.

Ce qui vaut pour le langage des sensations corporelles vaut également pour la conscience. À cause de la simplicité et de la spontanéité de sa ligne de type, l'expérience douloureuse ne se réduit pas à la perception confuse d'une stimulation nocive. Au contraire, la douleur est une entité essentiellement *bifaciale* : ce qui est présenté dans l'expérience douloureuse ne peut pas être identifié indépendamment de la stimulation nocive et de la réponse spontanée — ou au minimum de la *tendance* à répondre d'une certaine manière lorsque la ligne de type n'est pas effectivement exemplifiée. De même que le langage ne peut pas intervenir entre la douleur et son expression naturelle, la conscience corporelle ne peut pas dissocier les deux moitiés de la ligne de type sur laquelle elle est (normalement) fondée. Lorsque j'ai mal à l'épaule, la condition douloureuse dans laquelle se trouve une partie de mon corps m'apparaît comme simple dans la mesure où je ne peux pas clairement séparer, sur la seule base de mon expérience, la contribution de la stimulation nocive de celle de la réponse spontanée⁴⁴.

43. Voir Pears, *La pensée-Wittgenstein*, p. 300.

44. Je ne veux pas dire que les sensations corporelles soient les seules à exemplifier ce phénomène. Les « qualités » d'orientation (responsables du fait que je perçois un objet en haut, à droite, au fond, et ainsi de suite, dans mon champ perceptif) ressemblent aux sensations corporelles par la simplicité de leurs lignes de types. Il est intéressant de remarquer que les qualités d'orientation ne sont pas entièrement « externes » : le monde perçu n'est pas *intrinsèquement* orienté. Plutôt que d'introduire des modes de présentation perceptifs distincts des qualités de la scène perçue, on pourrait considérer les qualités d'orientation comme étant en partie constituées de réponses spontanées. Lorsque je perçois un objet à gauche, je ne perçois pas seulement cet objet, mais également ma propre tendance à tourner les yeux ou la tête vers la gauche (par exemple).

Selon la suggestion qui vient d'être émise, l'objet intentionnel de l'expérience de la douleur est plus qu'un simple dommage physique ; il inclut, ou du moins survient sur les réponses naturellement associées à celui-ci. Cette suggestion est non seulement plausible sur le plan conceptuel ; elle est également compatible avec les données phénoménologiques. Schilder a admirablement bien décrit ce trait fondamental de la douleur et ses conséquences au niveau de l'image du corps :

Il semble que toute douleur en un point de la surface du corps renforce le modèle postural en ce point, le distorde, avec des impressions d'enflure⁴⁵.

La partie du corps sur laquelle la douleur est ressentie draine toute l'attention. [...] Mais, en même temps, la partie douloureuse est isolée, quelque chose la pousse hors de l'image du corps⁴⁶.

Il est évident, à partir de cette description, que si l'expérience de la douleur est conçue comme une forme de perception objective, ce qui est perçu n'est pas un simple dommage physique. Le sujet qui souffre perçoit également et simultanément la réponse de son propre corps aux stimuli nocifs.

8. La réflexivité de la conscience objective interne

On remarquera à juste titre que la suggestion qui vient d'être émise n'explique en rien la différence pertinente entre la conscience externe et la conscience corporelle. On peut admettre que le même objet intentionnel — une qualité qui dépend de manière constitutive de relations entre des stimuli physiques et des réponses spontanées — est en principe accessible aux deux formes de conscience. Il est possible que la même chose me soit présentée à la conscience lorsque je ressens la douleur et lorsque je la vois chez autrui (ou chez moi-même dans un miroir). Je vais tenter de montrer que cette suggestion contient pourtant *in nuce* des éléments qui nous permettent de rendre compte de la spécificité de la conscience corporelle.

Nous avons vu que la TMS rend difficilement compte du lien intuitif entre la douleur et l'espace corporel. Considérons à présent une autre intuition ordinaire que nous avons sur la douleur. Il semble que la notion d'une douleur non ressentie soit incohérente : la douleur, contrairement au dommage physique, n'existe que par l'expérience que nous en avons. Pour la douleur, *esse est sentiri*. Certains matérialistes considèrent que le terme « douleur » ne désigne pas l'objet intentionnel de la douleur, et donc que cette intuition est compatible avec la TMS⁴⁷. À mon avis, ils se voilent la face devant un fait incontestable : la

45. Schilder, Paul, *L'image du corps*, Paris, Gallimard, 1968, p. 121.

46. *Ibid.*, p. 125.

47. Voir Tye, *Ten Problems of Consciousness*, pour qui la douleur est en fait l'expérience de la douleur. D'autres philosophes pensent que le concept de douleur est incohérent (Dennett, « Why You Can't Make a Computer that Feels Pain ») ou équivoque (Pitcher, « Pain Perception » ; Lycan, *Consciousness and Experience*).

douleur que je localise dans mon bras n'existe que parce que je la ressens ; elle ne résiste pas à l'anesthésie, et donc ne peut pas être un simple dommage physique.

Shoemaker fait observer qu'« au moins certaines formes de “comportement lié à la douleur” (*pain behavior*) — prendre de l'aspirine et appeler le médecin en sont de bons exemples — ne sont intelligibles *en tant que* comportement lié à la douleur que si l'on part du principe que le sujet est conscient (*aware*) de la douleur⁴⁸ ». Selon lui, il est impossible d'être « aveugle à soi-même » (*self-blind*) lorsque l'on a une douleur. L'existence de la douleur entraîne celle d'une expérience, même inattentive, qui porte sur elle.

L'observation de Shoemaker peut être transposée à l'ensemble des sensations corporelles. La forme générale de l'expérience corporelle est la suivante :

Expérience (cette partie corporelle est *F*)

Le prédicat *F* (par exemple, « avoir mal ») désigne une condition dans laquelle mon corps m'est présenté comme se trouvant. Même si cette condition m'apparaît typiquement comme simple, c'est un fait que mon corps ne peut s'y trouver qu'en présence de l'expérience dont elle est l'objet. L'expérience corporelle est essentiellement réflexive, et c'est parce qu'elle est réflexive que son objet intentionnel lui est lié de manière constitutive.

J'ai affirmé que la douleur, considérée comme l'objet intentionnel de la conscience corporelle, survient sur des relations naturelles entre des stimuli nocifs et des réponses spontanées de rejet. Si ces réponses se réduisaient à des gestes réflexes ou à des grimaces, la douleur ne dépendrait sans doute pas d'une expérience consciente. Mais elle serait également très éloignée de ce que nous appelons « douleur⁴⁹ ». La douleur humaine (comme peut-être certaines formes de douleur animale) a ceci de particulier que parmi les réponses qui la définissent, il y a précisément son entrée dans le champ de la conscience, qui en fait un objet susceptible d'attirer l'attention du sujet.

De l'impossibilité d'être aveugle à sa propre douleur, Shoemaker conclut que l'expérience de la douleur n'est pas une forme de perception objective. Elle serait plutôt une forme non perceptive d'introspection. Je pense que Shoemaker sous-estime la distinction entre « introspecter » une pensée ou une expérience et prendre conscience d'une sensation corporelle. Dans le premier cas, Shoemaker pourrait avoir raison lorsqu'il affirme que « la réalité connue et la faculté de la connaître sont pour ainsi dire faites l'une pour l'autre — l'une ne pourrait pas être ce qu'elle est sans l'autre⁵⁰ ». L'introspection n'est pas une forme de perception objective, avant tout parce que la notion d'illusion introspective est inintelligible. Il en va autrement de la conscience d'une sensation corporelle, à laquelle les traits présentés au commencement s'appliquent

48. Shoemaker, *The First-Person Perspective*, p. 228.

49. *Ibid.*, p. 229.

50. *Ibid.*, p. 245.

clairement : elle est localisée, elle peut attirer l'attention et, contrairement à la pensée introspective, elle peut être illusoire sous certains aspects. Plutôt que d'assimiler la conscience corporelle à l'introspection (ou inversement), je préfère la considérer comme une forme de conscience intermédiaire entre la perception externe, avec laquelle elle partage des traits qui définissent l'expérience objective, et l'introspection, qui n'a aucun de ces traits.

Selon mon hypothèse, que je présente ici comme une piste de recherche, c'est la réflexivité qui distingue (au moins partiellement) la conscience corporelle de la conscience externe. La douleur réelle est un état corporel essentiellement psychophysique ; elle suppose des stimuli physiques, mais elle engage de manière constitutive une expérience consciente. Lorsque je perçois la douleur « à l'extérieur », par exemple chez autrui, je perçois un état de choses complexe : un dommage ou un dérangement corporel et l'expérience du sujet focalisée sur le lieu douloureux. Mon acte de perception et l'expérience que je perçois sont des existences distinctes, car il n'y a qu'une relation causale entre la douleur et l'expérience que j'en ai. Par contraste, lorsque je ressens la douleur dans mon corps propre, un objet intentionnel du même type se présente à ma conscience, mais celle-ci est littéralement identique à l'expérience présentée ; il y a donc une relation constitutive entre la douleur et l'expérience que j'en ai. L'hypothèse est que cette différence contribue à expliquer pourquoi c'est seulement dans ce deuxième cas que je prends conscience de la douleur « de l'intérieur⁵¹ ».

De ce point de vue, l'intuition selon laquelle la douleur n'existe que par l'expérience que j'en ai est indissociable de l'intuition selon laquelle elle est nécessairement liée au corps propre. La douleur dépend ontologiquement de mon expérience parce que celle-ci est l'une des réponses qui la constitue. Aucun état du corps d'autrui ne pourrait avoir cette propriété. Si je suis réellement en présence d'une réponse consciente spontanée à des stimuli nocifs, et si je perçois cet état de choses « de l'intérieur », il ne peut concerner que mon propre corps, puisqu'il concerne ma propre expérience. Le critère (D) énoncé plus haut est donc essentiellement correct.

Une telle explication ne peut être considérée que comme une ébauche. Par exemple, il faudrait montrer plus en détail la manière dont la réflexivité de l'expérience corporelle est *apparente* au sujet, comment elle participe de l'effet que cela fait d'avoir une sensation corporelle. Je considère néanmoins qu'elle est plus prometteuse que la TMS, qui fait de l'objet intentionnel de l'expérience

51. On objectera que cette hypothèse ne vaut que si la conscience externe n'est pas (normalement) réflexive, alors que Searle, par exemple, a montré qu'elle l'était aussi (voir Searle, John, *L'intentionnalité*, traduction C. Pichevin, Paris, Minuit, 1987, ch. 2). L'interprétation de la théorie sui-référentielle de Searle est controversée (voir la réponse de Searle à Tyler Burge et à John McDowell dans Lepore, Ernest, et van Gulick, Robert, dir., *John Searle and His Critics*, Oxford, Blackwell, 1991), mais s'il veut dire que le contenu *apparent* de l'expérience visuelle inclut cette même expérience, il me semble que sa position est critiquable au moins sur des bases phénoménologiques.

de la douleur un simple dommage physique. Cette théorie est incapable de satisfaire aux intuitions récalcitrantes que nous avons sur la douleur et sur la conscience corporelle en général. L'explication qui vient d'être esquissée, en revanche, est en principe compatible avec le matérialisme, et a le mérite de prendre nos intuitions au sérieux. Si elle est sur la bonne voie, le matérialisme ne doit pas avoir peur des *qualia* corporels, du moins s'il ne craint pas ceux de la perception en général.

Fonctionnalisme et similarité phénoménale

Paul Bernier

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004902ar>
DOI : <https://doi.org/10.7202/004902ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Bernier, P. (2000). Fonctionnalisme et similarité phénoménale. *Philosophiques*, 27 (1), 99–114. <https://doi.org/10.7202/004902ar>

Résumé de l'article

Dans la foulée de divers arguments antiphysicalistes visant à montrer que les *qualia* ne sont pas fonctionnalisables, Ned Block a proposé un autre argument de ce type, qui repose sur son expérience de pensée de la Terre inversée. L'argument de Block montrerait qu'un sujet peut avoir deux expériences de couleur du même type « phénoménal » qui seraient de deux types fonctionnels distincts puisque, selon lui, elles auraient des contenus intentionnels distincts. Il existerait donc une différence fondamentale entre le contenu *phénoménal* des expériences sensorielles et leur contenu *intentionnel* et, dans la mesure où le contenu phénoménal ne pourrait être fonctionnalisé, le fonctionnalisme serait ainsi réfuté, à tout le moins en ce qui concerne les expériences sensorielles. Je soulève une objection qui, si elle est correcte, montre que la distinction que Block tente d'établir entre le contenu phénoménal et le contenu intentionnel des expériences est incohérente.

Fonctionnalisme et similarité phénoménale

PAUL BERNIER

bernierp@umoncton.ca
Université de Moncton

RÉSUMÉ. — Dans la foulée de divers arguments antiphysicalistes visant à montrer que les *qualia* ne sont pas fonctionnalisables, Ned Block a proposé un autre argument de ce type, qui repose sur son expérience de pensée de la Terre inversée. L'argument de Block montrerait qu'un sujet peut avoir deux expériences de couleur du même type « phénoménal » qui seraient de deux types fonctionnels distincts puisque, selon lui, elles auraient des contenus intentionnels distincts. Il existerait donc une différence fondamentale entre le contenu *phénoménal* des expériences sensorielles et leur contenu *intentionnel* et, dans la mesure où le contenu phénoménal ne pourrait être fonctionnalisé, le fonctionnalisme serait ainsi réfuté, à tout le moins en ce qui concerne les expériences sensorielles. Je soulève une objection qui, si elle est correcte, montre que la distinction que Block tente d'établir entre le contenu phénoménal et le contenu intentionnel des expériences est incohérente.

ABSTRACT. — Following the popular trend of antiphysicalist arguments which aim at showing that *qualia* cannot be functionalized, Ned Block has suggested yet another antiphysicalist argument, on the basis of his Inverted Earth thought experiment. The purpose of his argument is to show that a subject could have two color experiences which are of the same phenomenal type but of different functional types, insofar as they have different intentional contents. This argument would support the idea that there exists a substantial distinction between the *phenomenal* content of sensory experiences and their *intentional* content, and that functionalism is false, at least with respect to sensory experiences. I raise an objection to Block's argument which, if correct, shows that the distinction he proposes between the *phenomenal* and the *intentional* content of sensory experiences is incoherent.

1. Introduction

Selon la thèse fonctionnaliste, les états mentaux sont caractérisés par leurs rôles causaux relativement aux stimuli, aux autres états mentaux, et aux comportements. Pour un très grand nombre de philosophes de l'esprit la thèse fonctionnaliste constitue une solution physicaliste adéquate au problème cerveau-esprit, à tout le moins pour ce qui concerne les états mentaux intentionnels, comme les croyances et les désirs¹. Mais dans le cas de la conscience phénoménale, ou ce qu'on a appelé les « *qualia* », la situation est très différente. Comme le note Thomas Nagel au tout début de son célèbre article

1. Voir notamment Armstrong, D., *A Materialist Theory of Mind*, London, Routledge & Kegan Paul, 1968 ; Lewis, D., « An Argument for the Identity Theory », *Journal of Philosophy*, 67, 1970, p. 203-211 ; et Kim, J., « Émergence, modèles de réduction, et le mental », dans le présent numéro, p. 9-24.

« What Is It Like To Be a Bat? », « la conscience est ce qui rend le problème cerveau-esprit intraitable² ». Bien que cette idée de Nagel n'ait pas été universellement admise, dans la littérature récente, elle a eu une influence considérable, et un grand nombre de philosophes l'ont endossée sur la base de divers arguments³. Si l'un ou l'autre de ces arguments était effectivement fondé, cela montrerait que le fonctionnalisme n'est pas une solution généralement adéquate au problème cerveau-esprit, dans la mesure où les *qualia* résisteraient à une analyse fonctionnelle.

Mon objectif est de défendre le fonctionnalisme contre un argument qui a récemment été formulé par Ned Block. Comme je le précise dans ce qui suit, cet argument est une variation sur le thème du célèbre argument du spectre inversé. Mon objectif est donc plutôt modeste puisque si ma réfutation était correcte, elle aurait une portée relativement restreinte compte tenu de la diversité des arguments antiphysicalistes qui ont été suggérés pour montrer que les *qualia*, ou la conscience phénoménale, ne pourraient être fonctionnalisés. Mais cette discussion me paraît tout de même importante car si nous comprenons clairement pourquoi l'argument de Block est erroné, cela devrait nous placer dans une bonne position pour défendre le fonctionnalisme de façon générale et pour comprendre ce qui ne va pas avec les divers arguments visant à montrer qu'il serait impossible de fonctionnaliser les *qualia*.

2. L'inversion intrapersonnelle

Dans un article récent⁴, Ned Block a soutenu que même si nous devons accepter que les expériences sensorielles — par exemple l'expérience visuelle que j'ai lorsque je suis confronté visuellement à une tomate dans des conditions d'éclairage normales — ont un contenu intentionnel qui est fonctionnalisable,

2. Nagel, T., « What Is It Like To Be a Bat? », dans Block, N., Flanagan, O. et Güzelderde G., dir., *The Nature of Consciousness*, Cambridge (Mass.), MIT Press, 1997, p. 519 (paru initialement dans *Philosophical Review*, 83, 1974, p. 435-450).

3. Je pense principalement à l'argument des *qualia* absents, à l'argument de la connaissance de Frank Jackson, à l'argument du spectre inversé et à celui du fossé explicatif (*explanatory gap*). Voir respectivement Block, N., « Troubles With Functionalism », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980 ; Jackson, F., « Epiphenomenal Qualia », *Philosophical Quarterly*, 32, 1982, p. 127-136 ; Shoemaker, S., « The Inverted Spectrum », *Journal of Philosophy*, 74, 1981, p. 357-381 ; Levine, J., « Materialism and Qualia : The Explanatory Gap », *Pacific Philosophical Quarterly*, 64, 1983, p. 354-361. Il existe bien sûr des différences importantes dans la portée respective de ces arguments, mais tous visent à poser une difficulté de principe au physicalisme ou, de façon générale, au projet de naturalisation de l'esprit, et particulièrement à la thèse fonctionnaliste.

4. Block, N., « Inverted Earth », dans Tomberlin, J., dir., *Philosophical Perspectives*, vol. 4, Atascadero, Ridgeview, 1990, p. 31-52 ; réédité dans Block, Flanagan et Güzelderde, dir., *The Nature of Consciousness*, p. 677-693. Nos références ultérieures à ce texte renvoient à cette dernière édition.

cela ne suffirait pas à résoudre le problème de la conscience phénoménale puisque, selon lui, les expériences sensorielles auraient en outre un contenu purement phénoménal, ou qualitatif, qui ne pourrait être fonctionnalisé. Un des principaux objectifs de l'argument de Block est donc d'établir une distinction fondamentale entre le contenu *phénoménal* des expériences sensorielles et leur contenu *intentionnel*. Si mes objections dans ce qui suit sont correctes, alors cette distinction est trompeuse, et elle devrait être rejetée.

Dans un premier temps, il est utile de faire quelques remarques générales à propos de la principale motivation de l'argument de Block qui s'inscrit dans le cadre d'une discussion de l'argument du spectre inversé. En effet, Block croit que ce dernier, dans sa version *intrapersonnelle* standard qui a été suggérée par Sydney Shoemaker, serait sujet à certaines objections d'incohérence⁵. Même s'il n'est pas tout à fait clair qu'il tienne ces objections pour décisives, il suggère que son propre argument, soit l'argument de la Terre inversée, est préférable à celui du spectre inversé dans la mesure où, selon lui, ce type d'objections d'incohérence ne pourrait lui être adressé. Comme on le sait, l'argument du spectre inversé cherche à montrer qu'une individuation des expériences de couleur qui repose uniquement sur leur rôle fonctionnel ou sur les dispositions comportementales des sujets serait inadéquate dans la mesure où il serait possible d'imaginer que deux expériences de couleur puissent avoir des contenus phénoménaux différents alors qu'elles seraient du même type fonctionnel. Et, comme nous le verrons, l'argument de Block repose sur une expérience de pensée qui vise à établir que deux expériences de couleur pourraient avoir le même contenu phénoménal, alors qu'elles seraient de types fonctionnels différents.

Sydney Shoemaker a soutenu que si nous devons admettre la cohérence de l'argument du spectre inversé dans sa version *interpersonnelle*, c'est-à-dire l'idée que deux sujets puissent avoir des expériences phénoménalement inverses qui seraient du même type fonctionnel, alors celui-ci doit reposer sur la cohérence d'une version *intrapersonnelle* de cette idée. Autrement dit, la plausibilité de l'argument ne peut reposer uniquement sur des considérations relevant de l'attribution en troisième personne de prédicats comme « *x paraît bleu* » ou « *x paraît être de la même couleur que y* ». Selon Shoemaker, la cohérence d'une telle possibilité doit reposer sur le point de vue de la première personne. En effet, si la signification d'un prédicat comme « *x paraît bleu* » était uniquement déterminée par les dispositions linguistiques des sujets à l'appliquer dans différentes circonstances de stimulation, alors l'idée

5. Voir Block, « Inverted Earth », p. 681-682. Pour la version *intrapersonnelle* standard de l'argument du spectre inversé, voir Shoemaker, « The Inverted Spectrum ». Pour les dites « objections d'incohérence », voir Dennett, D., « Quining Qualia », dans Lycan, W., dir., *Mind and Cognition*, Oxford, Blackwell, 1990, p. 519-548 ; et Rey, G., *Mind Without Consciousness : A Discrepancy Between Explanatory and Moral Psychology*, Cambridge (Mass.), MIT Press, à paraître.

centrale de l'argument du spectre inversé serait inintelligible, à savoir par exemple que ce qui paraît bleu à Pierre paraît jaune à Jean (et inversement), alors que ce qui paraît bleu à Pierre produit chez lui un état mental du même type fonctionnel que celui produit chez Jean par ce qui lui paraît jaune. Car si c'était le cas, cela reviendrait à dire les deux choses suivantes, qui sont conjointement incohérentes : i) Pierre est disposé à dire « ceci paraît bleu » dans les circonstances où Jean est disposé à dire « ceci paraît jaune », et ii) dans ces circonstances, Jean et Pierre ont des états mentaux du même type fonctionnel. Et, bien évidemment, compte tenu de leurs dispositions à produire deux énoncés différents, leurs états mentaux respectifs devraient être catégorisés comme étant de deux types fonctionnels distincts.

L'idée d'une inversion *intrapersonnelle* peut être résumée de la façon suivante, en quatre étapes⁶. Premièrement, on imagine un sujet qui est tout à fait normal. À la deuxième étape, on suppose que des lentilles qui ont la propriété d'inverser le spectre des couleurs sont insérées sous la cornée du sujet qui aurait alors l'impression que les couleurs sont inversées. Par exemple, à cette étape, le sujet serait disposé à dire « Les tomates mûres me paraissent vertes » et « L'herbe me paraît rouge ». Évidemment, cela est insuffisant pour appuyer la thèse du spectre inversé puisque son expérience, disons de rouge, changerait du même coup de rôle fonctionnel. Mais l'idée est qu'après une certaine période d'adaptation pendant laquelle le sujet utiliserait les prédicats de couleur de façon confuse, il pourrait éventuellement les utiliser de façon conforme à l'usage intersubjectif admis dans sa communauté, et cela malgré l'inversion de son spectre. La troisième étape est cruciale puisque le sujet doit alors être en mesure de poser des jugements de similarité comme le suivant :

(1) Les tomates mûres me paraissent être de la même couleur que l'herbe me paraissait être jadis.

Le but de cette étape est de donner une plausibilité, sur la base d'intuitions en première personne, à l'idée que ce que cela fait au sujet que d'être confronté visuellement à des tomates mûres, maintenant, est différent de ce que cela lui faisait jadis, et par conséquent que le contenu de son expérience phénoménale est distinct de celui de l'expérience phénoménale d'un sujet « normal » qui se trouverait dans les mêmes conditions stimulatoires. Cependant, cela est encore insuffisant pour garantir la cohérence de la thèse du spectre inversé, dans la mesure où le sujet aurait encore des dispositions « déviantes », comme celle à énoncer (1), ce qui suffit pour nier que son expérience de couleur soit du même type fonctionnel que celui des expériences que *nous*, sujets normaux, aurions dans les mêmes conditions stimulatoires. Mais cette troisième étape est cruciale car elle en appelle à l'idée que le sujet jouit d'une certaine autorité épistémique le justifiant à croire et nous

6. Je reprends ici, pour l'essentiel, la présentation de Block dans « Inverted Earth », p. 681-682.

incitant à accepter, que ce que cela lui fait que d'être confronté visuellement à une tomate mûre maintenant est différent de ce que cela lui faisait jadis, même si cette étape est insuffisante pour établir l'identité fonctionnelle d'états mentaux phénoménalement distincts.

Finalement, à la quatrième étape, nous supposons que le sujet utilise les prédicats de couleur de façon tout à fait conforme à l'usage intersubjectif. Pour cela, c'est-à-dire pour éviter que le sujet ait la disposition « déviante » à produire des énoncés comme (1), nous devons supposer qu'il serait désormais affligé d'une amnésie d'un type très spécifique, en raison de laquelle il aurait oublié tout ce qui lui est arrivé avant l'insertion des lentilles inversantes. Donc, si nous acceptons que l'inversion a bel et bien eu lieu, c'est-à-dire si nous donnons foi aux jugements de similarité que le sujet était disposé à faire à la troisième étape avant que l'amnésie ne se produise, alors nous serions en présence d'une inversion *intrapersonnelle* du spectre, dans la mesure où rien ne nous contraint, logiquement, à accepter que, suite à l'amnésie, le spectre devrait nécessairement se ré-inverser à nouveau pour redevenir « normal ». Ce que cela lui fait que d'être visuellement confronté à des tomates mûres à la quatrième étape est différent de ce que cela lui faisait avant l'inversion des lentilles, alors que ces deux expériences auraient exactement le même rôle fonctionnel. La troisième étape met en évidence l'intuition en première personne selon laquelle ce que cela lui fait, à la quatrième étape, d'être confronté visuellement à des tomates mûres est différent de ce que cela nous fait. Et, l'hypothèse d'amnésie garantit que ses expériences et les nôtres sont bien du même type fonctionnel puisqu'à la quatrième étape le sujet ne serait plus disposé à faire des énoncés comme (1). Compte tenu de son amnésie, ses dispositions seraient conformes aux nôtres. C'est ainsi que la plausibilité d'une inversion *intrapersonnelle* permet de rendre intelligible l'inversion *interpersonnelle*.

Mais, comme Block le fait remarquer, certains philosophes se sont objectés à cet argument en proposant que le jugement de similarité en première personne qui est requis à la troisième étape est incohérent, ou bien parce que des jugements de ce type devraient être interprétés comme indiquant que le sujet utilise correctement les prédicats de couleurs mais qu'il a certains problèmes de mémoire, ou sinon que le sujet ne comprend tout simplement pas la signification des prédicats de couleur qu'il utiliserait de façon confuse. Il ne me paraît pas évident que ces objections soient dirimantes, et mon but ici n'est pas de les évaluer mais simplement de rappeler ce qui semble être la principale motivation de l'argument de Block. En effet, Block croit que ce type d'objection d'incohérence ne pourrait tout simplement pas être adressé à son propre argument. Et c'est précisément pour cette raison qu'il suggère que son argument doit être préféré à l'argument standard du spectre inversé. Mais si ma réfutation de l'argument de Block est correcte, c'est plutôt celui-ci qui est confronté à une objection d'incohérence. Cela dit, la question de savoir si les considérations que je propose pour réfuter l'argument de Block pourraient servir

de base à une objection de l'argument standard du spectre inversé est une question ouverte que je ne pourrai aborder dans cet article⁷.

3. L'argument de Terre inversée

L'argument de Block repose sur l'expérience de pensée suivante⁸. Supposons qu'il existe une planète, Terre inversée, qui diffère de la Terre uniquement des deux façons suivantes. Premièrement, les objets sur Terre inversée ont exactement les couleurs complémentaires de leurs contreparties sur la Terre. Par exemple, sur Terre inversée, les tomates mûres sont vertes, l'herbe est rouge, le ciel est jaune, etc. Block souligne que cette supposition repose sur une certaine conception réaliste des couleurs, à savoir que ce sont les objets eux-mêmes qui ont ces couleurs bizarres⁹. Deuxièmement, les habitants de Terre inversée qui sont les contreparties des gens qui parlent le français sur la Terre parlent ce qu'on pourrait appeler le français inversé. Le français inversé est homophonique avec le français, mais il en diffère uniquement en ce sens que les prédicats de couleur y sont inversés par rapport aux prédicats de couleur en français. C'est-à-dire que les locuteurs du français inversé traduiraient leur prédicat « rouge » par notre prédicat « vert » et ainsi de suite pour les autres prédicats de couleur. Alors que nous sommes disposés à appliquer le prédicat « rouge » aux tomates mûres sur Terre ainsi qu'à l'herbe sur Terre inversée, les locuteurs du français inversé seraient disposés à l'appliquer aux tomates mûres de Terre inversée et à l'herbe sur Terre. En somme, l'expérience de pensée repose sur la possibilité d'une double inversion, soit une inversion de la couleur des objets et une inversion de la dénotation des prédicats de couleur. Évidemment, ces deux inversions se trouvent en quelque sorte à s'annuler mutuellement, en ce sens que les locuteurs du français inversé accepteraient, sur Terre inversée, les énoncés suivants : « Le ciel est bleu », « L'herbe est verte » et ainsi de suite, tout comme nous les acceptons sur Terre.

7. Cette question se pose dans la mesure où l'argument standard du spectre inversé suggère également la possibilité que le sujet ait deux expériences d'un même type phénoménal mais de types fonctionnels différents, même si lorsqu'on parle du « spectre inversé » on entend généralement la possibilité inverse, à savoir celle de deux expériences de types phénoménaux distincts mais d'un même type fonctionnel. Je remercie un commentateur anonyme d'avoir attiré mon attention sur ce point. Une discussion élaborée de la première de ces possibilités, dans le cadre de l'argument standard du spectre inversé, irait au-delà des limites du présent article.

8. *Ibid.*, p. 682-685.

9. Comme il le note : « I mean everything *really* has these oddball colors. If you visited Inverted Earth along with a team of scientists from your university, you would all agree that on this planet, the sky is yellow, grass is red and so forth » (*ibid.*, p. 682). Une telle conception « réaliste » des couleurs, et l'interprétation que nous devons en donner dans le contexte de l'argument de Block, soulève certainement des questions intéressantes, mais je me permets de les ignorer et d'admettre cette hypothèse pour les fins de la discussion, même si elle demeure relativement vague et qu'elle porte à controverse.

À vrai dire, l'expérience de pensée proprement dite invoque une troisième inversion. Comme nous le verrons, l'argument vise à appuyer l'idée d'une inversion fonctionnelle *intrapersonnelle*, et elle repose sur le scénario suivant que Block nous demande d'imaginer en première personne¹⁰. Supposons qu'au temps t_1 je me trouve sur la Terre, disons que je suis étendu dans un hamac sur une plage de la Martinique, par une belle journée où le ciel est d'un bleu limpide. Contemplant la voûte céleste, j'aurais une certaine expérience de couleur, que j'appellerai E_1 , et je serais disposé à énoncer la phrase suivante : « Le ciel est bleu. » On suppose ensuite que je m'endors soudainement et qu'à ce moment une équipe de scientifiques entre en scène pour m'utiliser, à mon insu, comme sujet d'une expérience scientifique. On m'anesthésie et on m'insère sous la cornée des lentilles qui ont la propriété d'inverser le spectre des couleurs, de telle sorte que si je me réveillais sur Terre, alors le ciel me paraîtrait jaune, l'herbe rouge et ainsi de suite. Cependant, selon le scénario, je ne me réveille pas sur Terre car pendant mon sommeil les scientifiques auraient fait en sorte que je sois transporté sur Terre inversée, plus précisément dans le hamac où se serait trouvé la contrepartie de moi-même. Ainsi, lorsque je me réveillerais, tout me paraîtrait parfaitement normal : le ciel jaune de Terre inversée me paraîtrait bleu, et ainsi de suite, puisque les lentilles inversantes annuleraient en quelque sorte l'inversion *de re* qui existe entre les deux environnements. Donc au temps t_2 , je serais sur Terre inversée, sans le savoir, et contemplant la voûte céleste j'aurais une certaine expérience de couleur, que j'appellerai E_2 , et je serais encore une fois disposé à énoncer la phrase : « Le ciel est bleu », ainsi que la phrase : « Le ciel me paraît toujours aussi bleu qu'il me paraissait avant que je ne m'endorme. »

Avant d'aller plus loin, il importe de préciser la stratégie de Block dans son utilisation de cette expérience de pensée. Son but est d'arriver à donner foi à une instance du schéma de thèse suivant :

(T) Deux de mes expériences de couleur, disons E et E' , qui se produiraient respectivement sur la Terre et sur Terre inversée, auraient le même contenu phénoménal alors qu'elles seraient de types fonctionnels distincts.

La stratégie de Block consiste, premièrement, à suggérer un cas où il serait intuitivement contraignant de penser que deux expériences pertinentes, E et E' , seraient du même type phénoménal et, deuxièmement, à soutenir

10. Je modifie légèrement le scénario tel que formulé par Block, simplement dans le but de présenter mon argument d'une façon qui soit aussi claire que possible. Cela dit, ma formulation du scénario suit strictement la logique de celle de Block. Par ailleurs, en accord avec une suggestion de Block, je formule le scénario à la première personne, c'est-à-dire en me prenant moi-même comme sujet de l'expérience de pensée. Ce point est important puisque, comme nous le verrons, une des affirmations cruciales de l'argument est que deux expériences de couleur pertinentes seraient du même type phénoménal ; et cette affirmation doit reposer sur des intuitions que nous avons du point de vue de la première personne.

que dans ce cas nous avons de bonnes raisons de croire que E et E' seraient de types fonctionnels distincts. Mais pour bien comprendre son argument, il est important de voir qu'il repose sur les trois présupposés suivants :

(P1) Les expériences de couleur ont un contenu intentionnel.

(P2) Le contenu intentionnel d'une expérience de couleur est déterminé par la signification des prédicats de couleur dans la langue du sujet. (Par exemple, le contenu intentionnel de l'expérience E_1 serait déterminé par la signification du prédicat « bleu » en français, car E_1 me disposerait à produire des énoncés comme « Le ciel est bleu », vrais en français.

(P3) Le contenu intentionnel d'une expérience de couleur détermine son individuation fonctionnelle dans le sens suivant : si deux expériences de couleur ont le même contenu intentionnel (ou différents contenus intentionnels), alors elles sont du même type fonctionnel (ou de types fonctionnels différents).

Pour arriver à trouver deux expériences pertinentes qui permettent d'énoncer une instance vraie du schéma (T), Block doit développer davantage son expérience de pensée, car il est tout à fait douteux que les expériences E_1 et E_2 , telles que décrites plus haut, puissent servir à cette fin, compte tenu des présupposés (P2) et (P3) qui suggèrent que E_1 et E_2 seraient du même type fonctionnel, puisqu'il est tout à fait plausible qu'au temps t_2 mes prédicats de couleur aient la même signification qu'ils avaient à t_1 . L'expérience de pensée doit donc être développée de la façon suivante. Supposons qu'après le temps t_2 je reste sur Terre inversée pour une période de temps « assez longue ». Et supposons, de plus, que durant cette période, j'interagisse linguistiquement avec les habitants de Terre inversée, d'une façon qu'on pourrait qualifier de normale en ce sens que mes interactions linguistiques se conformeraient aux usages en vigueur, et qu'elles me permettraient de communiquer avec succès. Alors, selon Block, il serait tout à fait plausible qu'après cette période je ne parlais plus français, mais je parlais le français inversé¹¹. Ainsi, mes prédicats de couleur n'auraient plus une signification en français, mais une signification en français inversé. Ce point peut être illustré de la façon suivante. Supposons qu'au temps t_3 je sois resté sur Terre inversée pour une période de temps « assez longue », de sorte que mes prédicats de couleur aient changé de signification. Alors, si j'utilisais les prédicats « bleu », « jaune », et ainsi de suite, à t_3 , ceux-ci auraient dans ma bouche la signification qu'ils ont dans la bouche des locuteurs normaux de Terre inversée. Supposons de plus qu'à t_3 je me trouve, encore une fois, dans un hamac sur une plage par temps clair, dans des conditions aussi similaires que possibles à celles dans lesquelles j'ai eu

11. Un tel scénario correspond à ce qu'on a appelé des cas de transfert lent (*slow switching*), qui ont été discutés dans la littérature sur la question de la compatibilité entre l'externalisme sémantique et l'autorité de la première personne. Je crois que Tyler Burge fut le premier à suggérer cette idée, voir Burge, T., « Individualism and Self-Knowledge », *Journal of Philosophy*, 85, 1988, p. 652.

l'expérience E_1 . Et supposons, finalement, qu'à t_3 j'aie une expérience visuelle, disons E_3 , et que je sois alors disposé à dire : « Le ciel est bleu. » Dans la mesure où par « bleu » je signifierais ce que les habitants de Terre inversée signifient normalement lorsqu'ils utilisent ce prédicat, et compte tenu des présupposés (P2) et (P3), E_3 aurait un contenu intentionnel différent de celui de E_1 , et par conséquent ces deux expériences seraient de types fonctionnels distincts, alors qu'il est intuitivement contraignant de penser qu'elles auraient le même contenu phénoménal. C'est donc ainsi que Block suggère un exemple qui, selon lui, satisferait le schéma (T).

4. Un présupposé crucial de l'argument

Comme je l'ai déjà noté, cet argument est basé sur des intuitions que le lecteur devrait avoir du point de vue de la première personne. En effet, s'il semble intuitivement contraignant d'affirmer que ce que cela fait (*what it's like*) d'avoir l'expérience E_1 est similaire à ce que cela fait d'avoir l'expérience E_3 , c'est précisément parce que le lecteur à qui l'on présente l'argument doit être en mesure de concevoir que si lui-même se trouvait dans le scénario, tel que décrit pour le temps t_3 , alors il serait en mesure de faire un jugement de similarité vrai dans lequel il jugerait que ce que cela lui ferait d'avoir l'expérience E_3 serait phénoménalement similaire à ce que cela lui aurait fait d'avoir l'expérience E_1 . Autrement dit, me prenant moi-même comme sujet de l'expérience de pensée, je dois être en mesure de concevoir que si je me trouvais dans la situation contrefactuelle décrite, je serais en mesure de dire l'énoncé vrai suivant, au temps t_3 :

(S) Le ciel me paraît maintenant de la même couleur que dans le passé, il me paraît bleu.

Ce point semble difficilement contestable. Mais je crois qu'à partir du moment où nous tentons de clarifier les conditions sous lesquelles je pourrais effectivement produire un tel énoncé vrai, cela soulève un sérieux problème pour l'argument de Block, et plus précisément pour l'affirmation selon laquelle E_1 et E_3 auraient des contenus intentionnels distincts, et que ces expériences seraient de types fonctionnels distincts. Pour clarifier ce que j'ai à l'esprit, il est d'abord utile de reformuler ce point de façon un peu plus précise.

Pour les raisons que je viens de noter, je (en tant que sujet de l'expérience de pensée) dois être en mesure de concevoir que si je me trouvais dans les conditions décrivant ce qui se passe au temps t_3 , je pourrais exprimer correctement les deux jugements suivants :

(SP) Le ciel me paraît maintenant de la même couleur que dans le passé ;

(P) Le ciel me paraît bleu,

où « maintenant » réfère à t_3 et « dans le passé » devrait avoir dans son extension le temps t_1 , ou à tout le moins un certain temps t' auquel mes prédicats

de couleur avaient encore leur signification en français. S'il m'était impossible de concevoir que je puisse faire de tels jugements de similarité vrais — par exemple, si je devais imaginer qu'au temps t_3 j'aurais perdu toute trace mnémonique de ce qui me serait arrivé avant le moment (ou l'intervalle de temps) où mes prédicats de couleurs auraient changé de signification — alors je ne serais aucunement justifié, *du point de vue de la première personne*, à accepter que ce que cela me fait que d'avoir l'expérience E_3 serait phénoménalement similaire à ce que cela m'avait fait d'avoir l'expérience E_1 . Sinon, je pourrais tout aussi bien juger « correctement » que ce que cela me fait d'avoir l'expérience E_3 serait phénoménalement différent de ce que cela m'avait fait d'avoir E_1 . Ce point me semble difficilement contestable. Premièrement, il semble aller de soi que nous pouvons communément faire des jugements comme le suivant : « L'impression que la vision de ce mur me fait, relativement à la couleur, est la même que celle que la vision de cet autre mur m'a fait hier. » Mais s'il nous semble intuitivement plausible que nous puissions aisément faire de tels jugements, qui sont justifiés, cela semble tenir au fait que lorsque nous les faisons nous avons présent à l'esprit, sinon un souvenir exact de ce que nous avons vu, à tout le moins un certain *relatum* auquel nous comparons notre expérience présente. Car, sinon, ce serait une sérieuse erreur que de considérer de telles expressions comme exprimant un jugement : « *Anything would go!* ». Or, dans le cas qui nous intéresse, ce *relatum* doit être un souvenir de l'expérience E_1 et, dans la mesure où l'argument suppose que E_1 et E_3 ont le même contenu phénoménal, il doit s'agir d'un cas où ce souvenir serait fidèle, car si le *relatum* auquel je comparerais mon expérience présente était d'un type phénoménal distinct de celui de E_1 , alors de deux choses l'une : ou bien mon jugement de similarité serait vrai et le présupposé que E_1 et E_3 sont du même type phénoménal serait faux, ou bien ce présupposé serait vrai, et je ne serais pas disposé à faire un tel jugement de similarité, je serais plutôt disposé à nier ce jugement. Quoi qu'il en soit, l'argument nous oblige à accepter que le sujet devrait être en mesure d'énoncer (SP) et (P), au temps t_3 , et que ces énoncés soient vrais.

Mais si nous devons accepter cela, et si nous devons également accepter la raison offerte par Block pour maintenir que E_1 et E_3 sont de types fonctionnels distincts — à savoir qu'au temps t_3 le sujet ne s'exprimerait plus en français mais en français inversé — alors la question suivante se pose : quelle est la signification de ces phrases en français inversé (et en français)? Et, plus précisément, qu'est-ce que les prédicats « x me paraît de la même couleur que y » et « x me paraît bleu » signifient en français inversé (et en français)? J'aimerais proposer une analyse toute simple de ces prédicats, qui me semble tout à fait acceptable, mais qui a de fâcheuses conséquences pour l'argument de Block. Selon l'analyse que je vais suggérer, le prédicat relationnel « x me paraît de la même couleur que y » a, en français inversé, la même signification qu'en français, ce qui n'a rien de très surprenant. Mais, comme je l'explique dans la section suivante, sur la base de cette analyse du prédicat

relationnel, nous pouvons ensuite produire une analyse du prédicat monadique « *x* me paraît bleu » à l'aide d'une certaine propriété relationnelle, ce qui permet de soutenir que l'énoncé (P), que je serais disposé à énoncer si j'étais dans les conditions décrites au temps t_3 , doit être interprété en français, et non pas en français inversé comme le requiert l'argument de Block.

5. Pourquoi l'argument doit être rejeté

Compte tenu de l'expérience de pensée telle que décrite, le prédicat relationnel « *x* me paraît de la même couleur que *y* », en français inversé, doit avoir dans son extension, à tout le moins, des paires d'objets que les habitants de Terre inversée s'accordent à catégoriser comme étant de la même couleur lorsqu'ils sont observés dans des conditions normales d'éclairage, par exemple des paires d'objets comme les suivantes : $\langle \text{herbe}_{TI}, \text{brocoli}_{TI} \rangle$, $\langle \text{citron-1}_{TI}, \text{citron-2}_{TI} \rangle$, $\langle \text{ciel}_{TI}, \text{mer}_{TI} \rangle$, etc. (où l'indice « *TI* » rend explicite qu'il s'agit d'objets de la Terre inversée). De la même manière, des paires d'objets que *nous* nous accordons à catégoriser comme étant de la même couleur tombent dans l'extension de ce prédicat en français. Si ces deux prédicats avaient dans leur extension uniquement des paires d'objets de Terre inversée et des paires d'objets de la Terre, respectivement, on pourrait dire qu'ils ont des significations différentes, puisqu'ils auraient manifestement des extensions différentes. Mais si c'était le cas, alors le jugement de similarité (SP) ne pourrait être vrai ni en français ni en français inversé, puisqu'une application correcte de ce prédicat, par moi dans l'expérience de pensée, requiert que le prédicat ait également dans son extension des paires d'objets « hétéroclites », c'est-à-dire des paires dont un des deux membres est un objet de la Terre et l'autre un objet de Terre inversée. Ainsi, en plus des paires d'objets notées plus haut, le prédicat, en français inversé, devrait également avoir dans son extension des paires d'objets comme les suivantes : $\langle \text{ciel}_{TI}, \text{citron}_T \rangle$, $\langle \text{herbe}_{TI}, \text{tomate}_T \rangle$, $\langle \text{citron}_{TI}, \text{mer}_T \rangle$, etc. Et, bien sûr, ces paires d'objets « hétéroclites » devraient également être dans l'extension du prédicat en français. Par conséquent, le prédicat aurait exactement la même extension en français et en français inversé. Cela peut être illustré simplement en imaginant que si les habitants des deux planètes pouvaient voyager régulièrement d'une planète à l'autre, et s'ils faisaient des jugements à propos de la similarité des couleurs de n'importe quelle paire d'objets dans le domaine étendu comprenant les objets des deux planètes, alors leurs jugements de similarité seraient en parfait accord. Ce qui bien sûr ne veut pas dire que leurs attributions de couleurs spécifiques à des objets s'accorderaient. Tout au contraire. Autrement dit, le prédicat monadique « *x* me paraît bleu » n'aurait pas la même signification dans les deux langues, puisque confronté à un citron_T n'importe quel locuteur normal du français serait disposé à utiliser le prédicat « jaune » alors qu'un locuteur normal du français inversé serait disposé à utiliser le prédicat « bleu ».

Cette analyse du prédicat relationnel peut sembler anodine, dans la mesure où elle ne fait que rendre explicite les présupposés sémantiques de l'expérience de pensée, mais elle pose une difficulté pour l'argument de Block. En effet, même si le prédicat relationnel a certaines paires d'objets « hétéroclites » dans son extension, cela ne permet pas de montrer comment le sujet pourrait faire un énoncé vrai en disant (SP) au temps t_3 car, au contraire, selon cette analyse l'énoncé (SP) serait manifestement faux, qu'il soit interprété en français ou en français inversé, ce qui bien sûr ne fait pas de différence. Cela paraît évident puisque $\langle \text{ciel}_{TI}, \text{ciel}_T \rangle$ n'est manifestement pas dans l'extension du prédicat. Cette analyse est donc insuffisante, puisqu'il est intuitivement contraignant de penser que le sujet pourrait effectivement exprimer un jugement vrai en énonçant (SP) au temps t_3 . À mon avis, la seule façon de montrer comment cela est possible est tout simplement de modifier l'analyse du prédicat relationnel d'une manière qui tienne compte de la possibilité que les expériences jugées similaires soient produites par la médiation d'un mécanisme d'inversion des couleurs, comme les lentilles inversantes supposées dans l'expérience de pensée. En fait, c'est précisément ce qui semble être à l'œuvre dans notre compréhension du scénario de l'expérience de pensée. L'analyse modifiée est la suivante (notons qu'il est inoffensif de supposer que le prédicat est défini uniquement pour les conditions d'éclairage normales) :

(R) « x me paraît de la même couleur que y » est satisfait, en français inversé, *si et seulement si* :

i) ni mon expérience de x , $E(x)$, ni mon expérience de y , $E(y)$, n'ont été produites via un mécanisme d'inversion des couleurs, ou toutes les deux l'ont été, et

$\langle x, y \rangle = \{ \langle \text{herbe}_{TI}, \text{brocoli}_{TI} \rangle, \langle \text{mer}_{TI}, \text{ciel}_{TI} \rangle, \langle \text{herbe}_T, \text{tomate}_{TI} \rangle, \langle \text{ciel}_{TI}, \text{citron}_T \rangle, \dots \}$

ii) une seule de mes expériences $E(x)$ et $E(y)$ a été produite via un mécanisme d'inversion des couleurs, et

$\langle x, y \rangle = \{ \langle \text{herbe}_{TI}, \text{tomate}_{TI} \rangle, \langle \text{citron}_{TI}, \text{ciel}_{TI} \rangle, \langle \text{ciel}_{TI}, \text{ciel}_T \rangle, \langle \text{citron}_{TI}, \text{citron}_T \rangle, \dots \}$

Cette analyse n'a rien de très surprenant. Elle est même plutôt banale, puisqu'elle ne fait que rendre explicite les intuitions sémantiques de l'expérience de pensée. Aussi, il n'est pas étonnant que, selon cette analyse, si j'énonçais (SP) au temps t_3 , cet énoncé serait vrai, et cela sans que nous ayons à supposer ou à nier que j'énoncerais (SP) en français ou en français inversé, puisque l'analyse du prédicat relationnel en français est exactement la même. Mais si nous nous en tenons uniquement à cette analyse du prédicat relationnel — c'est-à-dire si nous faisons abstraction du présupposé (P2) —, elle est insuffisante pour affirmer que les expériences E_1 et E_3 ont des contenus intentionnels distincts. En fait, cette analyse et la disposition du sujet à énoncer (SP) suggère plutôt qu'elles auraient le même contenu intentionnel, si

nous supposons par exemple que deux expériences de couleur ont le même contenu intentionnel si, et seulement si, le sujet est disposé à accepter des jugements les comparant, comme (SP). Et cette suggestion est certainement tout aussi acceptable que le présupposé (P2). Il semble donc qu'à ce stade, Block devrait nous donner des raisons de préférer (P2). À mon avis, cela suffit à montrer que l'argument de Block n'est pas du tout contraignant. Mais ce n'est pas tout.

Le problème plus sérieux pour l'argument de Block est de savoir comment (P) devrait être interprété ou, plus précisément, comment le prédicat monadique « x me paraît bleu », tel que je l'utiliserais au temps t_3 , doit l'être. Or, il semble qu'il existe une façon tout à fait naturelle de le faire sur la base de (R). (Ici aussi, il est inoffensif de supposer que le prédicat est défini uniquement pour les conditions d'éclairage normales.) L'analyse est la suivante :

(M_{FI}) « x me paraît bleu » est satisfait, en français inversé, *si et seulement si* :

x me paraît être de la même couleur que le ciel_{TI} si mon expérience du ciel_{TI}, ou $E(\text{ciel}_{TI})$, est produite d'une façon qui n'implique pas un mécanisme d'inversion des couleurs, ou de la même couleur que le citron_{TI} si mon expérience $E(\text{citron}_{TI})$ est produite d'une façon qui implique un mécanisme d'inversion des couleurs.

(M_F) « x me paraît bleu » est satisfait, en français, *si et seulement si* :

x me paraît être de la même couleur que le ciel_T si mon expérience $E(\text{ciel}_T)$ est produite d'une façon qui n'implique pas un mécanisme d'inversion des couleurs, ou de la même couleur que le citron_T si mon expérience $E(\text{citron}_T)$ est produite d'une façon qui implique un mécanisme d'inversion des couleurs.

La première chose importante à noter est que cette analyse repose essentiellement sur le prédicat relationnel tel que défini par (R), mais aussi sur la sélection d'exemplaires pertinents d'expériences auxquels l'expérience définie, soit mon expérience de bleu, est comparée. Deuxièmement, il est utile de se servir des expressions « le ciel_{TI} », « le citron_{TI} », « le ciel_T », et « le citron_T » pour formuler simplement ces conditions, mais rien d'essentiel ne repose sur cette façon de procéder. D'autres termes singuliers feraient tout aussi bien l'affaire. En fait, il serait peut-être même préférable de penser que les exemplaires pertinents doivent être sélectionnés de façon démonstrative. La seule contrainte est que ces termes singuliers sélectionnent les exemplaires de telle sorte que « x me paraît bleu » signifie en français inversé ce que « x me paraît jaune » signifie en français, et inversement¹². Finalement, il semble

12. Cette contrainte sur le choix des termes singuliers servant à sélectionner les exemplaires pertinents peut être précisée de la façon suivante. Supposons que a_1 , a_2 , a_3 , a_4 sont les termes qui jouent respectivement les rôles de « le ciel_{TI} », « le citron_{TI} », « le ciel_T », et « le citron_T » dans les conditions (M_{FI}) et (M_F). Ces termes doivent être tels que des locuteurs normaux du français et du français inversé s'accorderaient à dire que le prédicat « x est de la même couleur que y » i) est satisfait par $\langle a_1, a_4 \rangle$ et $\langle a_2, a_3 \rangle$ (et symétriquement), et ii) qu'il n'est satisfait par aucune des paires suivantes : $\langle a_1, a_2 \rangle$, $\langle a_3, a_4 \rangle$, $\langle a_1, a_3 \rangle$, $\langle a_2, a_4 \rangle$ (et symétriquement).

évident que, selon cette analyse, « x me paraît bleu » a des extensions différentes en français et en français inversé.

Cette analyse semble tout à fait acceptable. Encore une fois, elle ne fait que rendre explicite les intuitions sémantiques présupposées dans notre compréhension de l'expérience de pensée¹³. Mais elle semble être en conflit avec la raison que Block propose pour appuyer l'idée que E_1 et E_3 auraient des contenus intentionnels différents, à savoir qu'au temps t_3 le sujet s'exprimerait en français inversé et non pas en français. La raison de ce conflit est que si nous acceptons cette analyse, et si nous acceptons également qu'au temps t_3 le sujet doit pouvoir énoncer (P) et que cet énoncé serait vrai, alors il semble que le sujet devrait s'exprimer en français, et non pas en français inversé. Supposons que j'énonce (P) au temps t_3 . Alors cet énoncé pourrait difficilement impliquer que le ciel_{TI}, tel que je l'observerais au temps t_3 , me paraîtrait de la même couleur que le ciel_{TI} me paraîtrait si mon expérience $E(\text{ciel}_{TI})$ était produite d'une façon qui n'implique *pas* un mécanisme d'inversion des couleurs. Et, il n'impliquerait pas plus que le ciel_{TI}, tel que je l'observerais au temps t_3 , me paraîtrait de la même couleur que le citron_{TI} si mon expérience $E(\text{citron}_{TI})$ était produite d'une façon qui implique un mécanisme d'inversion des couleurs. Mais, si j'énonçais (P) en français inversé, il découle logiquement que l'une ou l'autre de ces conséquences devrait être vraie.

Le point est tout simplement que s'il nous semble intuitivement contraignant d'accepter l'idée que E_1 et E_3 ont le même contenu phénoménal, c'est précisément parce que nous comprenons que ces deux expériences disposeraient le sujet à produire deux énoncés vrais de « Le ciel me paraît bleu » (ou « Ceci me paraît bleu ») dans lesquels il aurait l'intention d'utiliser « me paraît bleu » dans *un seul et même sens*! Et ce sens ne peut être que celui que nous donnons à ce prédicat en français. Ainsi, il est absurde de dire qu'au temps t_3 le sujet utilise ce prédicat monadique en français inversé, du moins si celui-ci doit être entendu dans le sens que j'ai proposé, ce qui me semble difficilement contestable puisque mon analyse n'est qu'une explicitation des intuitions sémantiques qui nous permettent de comprendre l'expérience de pensée. De plus, si l'énoncé (P) est interprété comme étant en français, en accord avec l'analyse notée plus haut, il devient non seulement cohérent mais vrai. Car, en effet, cet énoncé semble bien impliquer que le ciel_{TI}, tel que je l'observerais au temps t_3 , me paraîtrait de la même couleur que le ciel_T si mon expérience $E(\text{ciel}_T)$ était produite d'une façon qui n'implique pas un mécanisme d'inversion des couleurs, et que le ciel_{TI}, tel que je l'observerais

13. Cette analyse semble être en accord avec l'esprit de l'analyse des prédicats de couleur proposée par Wilfrid Sellars dans son article « Empiricism and the Philosophy of Mind », dans Feigl, H. et Scriven M., dir., *Minnesota Studies in Philosophy of Science*, vol. 1, Minneapolis, University of Minnesota Press, 1956, récemment réédité dans Sellars, W., *Empiricism and the Philosophy of Mind*, Cambridge (Mass.), Harvard University Press, 1997.

au temps t_3 , me paraîtrait de la même couleur que le citron_T si mon expérience $E(\text{citron}_{TI})$ était produite d'une façon qui implique un mécanisme d'inversion des couleurs. C'est pourquoi nous devons rejeter l'affirmation, cruciale à l'argument de Block, selon laquelle les expériences E_1 et E_3 auraient des contenus intentionnels différents, et donc qu'elles seraient de types fonctionnels distincts. Mon argument suggère que ces expériences auraient bel et bien le même contenu intentionnel et qu'elles seraient du même type fonctionnel.

Avant de conclure, il me paraît utile de reformuler mon argument de façon un peu plus systématique, et cela sur la base d'un principe général qui rend explicite une relation conceptuelle importante entre le prédicat relationnel « x me paraît de la même couleur que y » et le prédicat monadique « x me paraît bleu ». Ce principe systématise tout simplement l'idée suivante. Quelles que soient les paires d'objets qui satisfont le prédicat relationnel dans une langue L , les objets dans ces paires devraient être tels que si l'un d'eux satisfait le prédicat monadique, l'autre doit le satisfaire également. (Dans ce qui suit j'utilise « Sxy » pour représenter le prédicat relationnel « x me paraît de la même couleur que y », et « Bx » pour représenter le prédicat monadique « x me paraît bleu ».) Le principe général peut être formulé de la façon suivante :

$$(PG) \forall(x) \forall(y) (Sxy \supset (Bx \equiv By))$$

Ce principe semble très plausible et difficile à nier. Et, comme je l'ai proposé, il semble que nous devons admettre qu'au temps t_3 , le sujet serait en mesure d'énoncer (SP) et (P) et que ces énoncés seraient vrais. Sur cette base nous pouvons formuler le raisonnement suivant :

- 1) $B\text{ciel}_{TI}$ de (P)
- 2) $S(\text{ciel}_{TI}, \text{ciel}_T)$ de (SP)
- 3) $B\text{ciel}_T$ de 1, 2, et (PG)
- 4) $B\text{ciel}_{TI} \ \& \ B\text{ciel}_T$ de 1 et 3

Or, dans la mesure où le principe (PG) est intuitivement contraignant, il semble que ce soit là un raisonnement que le sujet de l'expérience de pensée pourrait facilement faire au temps t_3 . Mais il est évident que si la ligne 4 était interprétée en français inversé, selon l'analyse que nous avons suggérée plus haut, cet énoncé serait incohérent. Mais il est parfaitement cohérent s'il est interprété en français, en accord avec (M_T).

Un fonctionnaliste semble donc être dans une excellente position pour rejeter la raison offerte par Block à l'appui de l'idée que nous pourrions facilement imaginer une instance vraie du schéma (T), puisqu'il ne nous a donné aucune bonne raison de croire que les deux expériences pertinentes, soit E_1 et E_3 , seraient de types fonctionnels différents. De plus, bien que mon argument présuppose (P1) et (P3), les considérations que j'ai évoquées suggèrent que le présupposé (P2) serait trompeur, dans la mesure où il semble que ce ne soit

pas tant la signification des prédicats de couleur qui détermine le contenu intentionnel des expériences de couleurs, ou à tout le moins pas directement, mais que celui-ci est déterminé plutôt par la signification de prédicats ayant la forme « x me paraît C » (où « C » doit être remplacé par un prédicat de couleur).

Selon Block, son argument permettrait d'établir l'existence d'une distinction substantielle entre le contenu intentionnel des expériences sensorielles et leur contenu purement phénoménal. Mais si mon argument est correct, cette distinction est trompeuse et nous devons la rejeter. Comme je l'ai noté dans la première section, la principale motivation de Block est de proposer un argument qui ne serait pas confronté à certaines supposées objections d'incohérence auxquelles l'argument du spectre inversé serait confronté, dans sa version *intrapersonnelle* classique. Bien qu'il ne me paraisse pas évident que ces objections d'incohérence soient fondées, comme je l'ai noté plus haut, et bien que mon objectif ici n'est pas de les évaluer, il me semble qu'il serait utile de discuter cette question à la lumière des considérations que j'ai proposées. Mais c'est là une tâche que je dois laisser pour une autre occasion. Quoi qu'il en soit, une chose est claire : si mon argument est correct, alors il n'y a aucune raison de penser que l'argument de la Terre inversée devrait être préféré à l'argument classique du spectre inversé, que ce dernier soit ou non confronté à des objections d'incohérence¹⁴.

14. Je remercie Renée Bilodeau, Murray Clarke, Nicholas Kaufmann, Daniel Laurier, Brian McLaughlin, Kai Nielsen, Claude Panaccio, et Richard Vallée pour leurs commentaires sur des versions antérieures de ce texte.

Ce que pense un individu peut-il expliquer ce qu'il fait?

Pierre Jacob

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004936ar>
DOI : <https://doi.org/10.7202/004936ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Jacob, P. (2000). Ce que pense un individu peut-il expliquer ce qu'il fait?
Philosophiques, 27 (1), 115–138. <https://doi.org/10.7202/004936ar>

Résumé de l'article

Un partisan du monisme matérialiste qui souscrit au réalisme intentionnel s'engage à octroyer aux contenus (ou aux propriétés sémantiques) des représentations mentales une efficacité causale dans la production d'effets physiques. Deux difficultés se présentent : d'une part, le caractère extrinsèque des propriétés sémantiques paraît inconciliable avec leur efficacité causale (c'est le défi de l'externalisme) ; d'autre part, comme un matérialiste suppose que toute entité qui exemplifie une propriété sémantique exemplifie simultanément des propriétés physiques, celles-ci risquent de priver celle-là d'efficacité causale (c'est la menace de l'exclusion explicative). La première section est consacrée au premier problème, et la seconde section au second problème.

Ce que pense un individu peut-il expliquer ce qu'il fait?'

PIERRE JACOB

jacob@poly.polytechnique.fr

Institut des sciences cognitives

RÉSUMÉ. — Un partisan du monisme matérialiste qui souscrit au réalisme intentionnel s'engage à octroyer aux contenus (ou aux propriétés sémantiques) des représentations mentales une efficacité causale dans la production d'effets physiques. Deux difficultés se présentent : d'une part, le caractère extrinsèque des propriétés sémantiques paraît inconciliable avec leur efficacité causale (c'est le défi de l'externalisme) ; d'autre part, comme un matérialiste suppose que toute entité qui exemplifie une propriété sémantique exemplifie simultanément des propriétés physiques, celles-ci risquent de priver celle-là d'efficacité causale (c'est la menace de l'exclusion explicative). La première section est consacrée au premier problème, et la seconde section au second problème.

ABSTRACT. — A philosopher who subscribes jointly to physicalism and to intentional realism is bound to grant content (or semantic properties) causal efficacy in the causation of physical effects. He must then overcome two problems. First, the extrinsicness of semantic properties seems to deprive them of causal power (the externalist threat). Secondly, on minimalist physicalist assumptions, no entity can instantiate a semantic property unless it simultaneously instantiates some physical properties. A representation's semantic property risks therefore to be deprived of causal efficacy by its physical properties (the problem of causal-explanatory exclusion). I deal with the first problem in the first section of this paper and with the second problem in the second section.

1. Introduction

La psychologie partage avec les autres sciences (dont la cosmologie, l'astronomie, la chimie, la biologie moléculaire ou la géologie) l'ambition de fournir des explications causales. On soutient généralement que les relations causales unissent des événements : l'événement *c* produit l'événement *e*². Dans les explications non psychologiques, on dit volontiers qu'un court-circuit a provoqué un incendie ou que la collision entre une météorite et la Terre a provoqué l'extinction des dinosaures. Toute explication n'est pas causale³.

1. Merci à deux lecteurs anonymes de la revue pour leurs remarques.

2. Soit que comme Davidson, 1967, on tienne la relation causale pour une relation purement extensionnelle entre deux individus, soit que comme Kim, 1973, on la tienne pour une relation intensionnelle entre deux exemplifications de propriétés à deux instants. Dans la suite, je supposerai qu'on peut faire référence à une cause et à un effet — les *relata* de la relation causale — sans mentionner aucune de leurs propriétés respectives. Mais je supposerai que dans une explication causale, on ne peut mentionner les faits constitutifs de l'*explanans* et de l'*explanandum* sans faire référence à certaines propriétés de la cause et de l'effet.

3. On peut, par exemple, tenir la démonstration d'un théorème mathématique à partir d'axiomes pour une explication de la vérité du théorème. On peut également tenir la traduction

Mais toute explication causale ou non causale relie des faits. Pour qu'une explication soit causale, il faut, d'une part, qu'il existe une relation causale entre l'événement mentionné dans l'*explanandum* (la cause *c*) et l'événement mentionné dans l'*explanans* (l'événement *e*). Il faut, d'autre part, que le fait que *c* exemplifie la propriété *F* soit causalement responsable du fait que *e* exemplifie la propriété *G*. Nombre de philosophes supposent de surcroît que le fait que *c* est *F* ne peut être causalement responsable du fait que *e* est *G* à moins qu'il existe une corrélation nomique entre les exemplifications de la propriété *F* et les exemplifications de la propriété *G*⁴.

À la différence d'une explication non psychologique, une explication psychologique fait souvent (sinon toujours) référence au contenu (ou à la propriété sémantique) d'au moins une représentation mentale d'un individu⁵. Dans le domaine psychologique, les relations causales pertinentes tiennent tantôt entre deux représentations mentales — lorsqu'une pensée produit une autre pensée —, tantôt entre une représentation mentale et un geste corporel.

Il existe une distinction entre ce qui arrive à un individu et ce qu'il fait (c'est-à-dire son comportement) : si un kinésithérapeute prend ma tête entre ses mains et lui fait subir une rotation de 90° vers la droite, le mouvement de ma tête est quelque chose qui m'arrive. Si j'effectue spontanément une rotation de la tête de 90° vers la droite, ce mouvement de ma tête est quelque chose que je fais. Cette distinction se fonde sur le contraste entre une cause interne et une cause externe du mouvement corporel : dans le second cas, mais non dans le premier, la cause de la rotation de ma tête est dans mon cerveau. Parmi les comportements d'un individu, certains sont intentionnels et d'autres ne le sont pas : sont intentionnels les comportements impliquant des gestes corporels guidés par des états internes doués de contenus représentationnels. Lorsque je ronfle et lorsque je vomis, j'émetts des sons et je contracte mes muscles abdominaux. Mais ni les sons que j'émetts ni mes contractions abdominales n'ont pour causes des représentations internes douées de contenus ou de propriétés sémantiques. Un comportement intentionnel peut être volontaire ou involontaire selon que le contenu de la représentation mentale qui guide le geste corporel est conceptuel ou non conceptuel : un geste corporel guidé par le contenu conceptuel d'une croyance fait partie d'un comportement intentionnel volontaire ; un geste corporel guidé par le contenu

d'une expression *e* d'une langue *L* par une expression *e'* appartenant à une langue *L'* différente de *L* pour une explication de la signification de la première expression. Dans ce cas, la démonstration du théorème et la traduction seront des explications non causales.

4. La corrélation nomique sera généralement supposée admettre des exceptions.

5. Je suppose que les croyances d'un individu sont des représentations mentales. Toute représentation n'est pas mentale : un énoncé, une photographie ou une carte géographique sont des représentations non mentales. Je suppose qu'une représentation mentale est un état physique du cerveau d'un individu. À la différence d'une représentation non mentale, une représentation mentale a des propriétés neurophysiologiques.

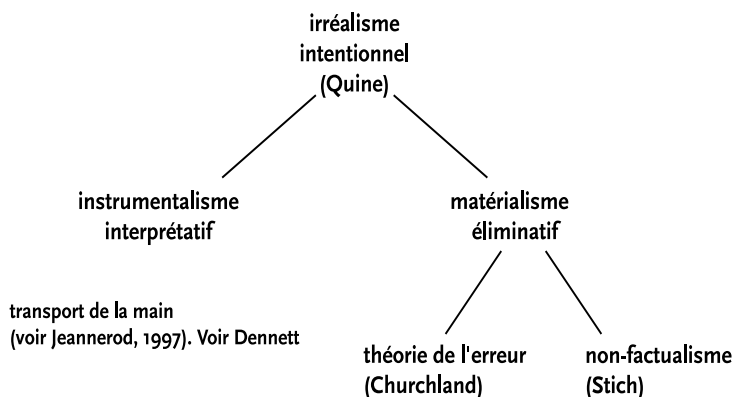
non conceptuel d'une représentation sensorielle fait partie d'un comportement intentionnel non volontaire⁶.

Imaginons que je transporte ma main droite à quarante centimètres de l'axe de mon buste et que, pendant le transport, j'ajuste (non consciemment) l'écart entre mon pouce et mon majeur au diamètre d'un verre contenant de l'eau ; je saisis le verre et bois une gorgée d'eau. La séquence ordonnée des gestes moteurs de mon bras et de ma main fait partie d'un comportement intentionnel volontaire⁷ : mes gestes ont pour causes ma croyance que le verre placé devant moi contient de l'eau et mon désir d'une gorgée d'eau.

Supposons que ma croyance *c* — une représentation mentale — est un état physique de mon cerveau (ou de mon système nerveux central) : l'exemplaire particulier de ma croyance qui a contribué à produire causalement la série de mes gestes corporels *m* n'est autre qu'un état concret de mon cerveau. Ma croyance *c* exemplifie donc plusieurs propriétés non sémantiques — des propriétés physiques, chimiques et neurophysiologiques. Ma croyance *c* a aussi un contenu (ou une propriété sémantique) : elle représente le fait que le verre placé devant moi contient de l'eau. La propriété sémantique de ma croyance possède-t-elle une efficacité causale dans le processus au cours duquel j'ai accompli la série de gestes corporels que j'ai décrits? Un matérialiste peut incontestablement tenir une représentation mentale pour la cause d'effets physiques grâce à l'identification entre un exemplaire concret de la représentation mentale et un état cérébral concret de l'individu. Mais pour assurer le transport de ma main droite à quarante centimètres de l'axe de mon buste et pour refermer la pince de mes doigts efficacement sur deux points opposés de la surface du verre, d'innombrables signaux électriques et chimiques ont dû être envoyés par mon cerveau le long des axones et des fibres nerveuses de mon

6. Malgré les critiques qui lui ont été adressées par McDowell, 1994, j'admets la distinction entre le contenu conceptuel et le contenu non conceptuel d'une représentation. La typologie des comportements de ce paragraphe est fidèle à la conception développée par Dretske, 1988, à ceci près que Dretske, 1988, ne reconnaît pas la distinction entre les comportements intentionnels volontaires et involontaires. Pour lui, un geste corporel produit par une représentation sensorielle douée d'un contenu non conceptuel fait partie d'un comportement instinctif non intentionnel. Selon lui, lorsqu'une représentation sensorielle contribue à la production d'un geste corporel, et quoique la représentation soit douée d'un contenu non conceptuel, ce contenu est néanmoins dépourvu d'efficacité causale dans la production du geste. Dans le chapitre VIII de Jacob, 1997, et dans Jacob, 1998, j'ai expliqué en détail pourquoi, contrairement à Dretske, j'admets que le contenu non conceptuel d'une représentation sensorielle peut être causalement efficace dans la production d'un comportement intentionnel non volontaire.

7. Ce qui ne veut pas dire que chaque étape motrice est planifiée consciemment. Les neurosciences cognitives montrent notamment que, dans la saisie d'un objet perçu visuellement, l'écart entre le pouce et l'index est calculé automatiquement pendant le transport de la main (voir Jeannerod, 1997).



système nerveux central aux muscles de ma main et de mon bras droits afin de commander leur contraction et leur décontraction successives.

La plupart des philosophes matérialistes qui souscrivent à ce qu'on peut appeler l'« irréalisme intentionnel » sont enclins à conférer aux propriétés non sémantiques (physiques, chimiques et neurophysiologiques) de mon état cérébral une efficacité causale dans ce processus et ils sont prêts à renoncer à la thèse de l'efficacité causale des propriétés sémantiques des représentations mentales. On peut représenter ainsi les principales versions de l'irréalisme intentionnel :

Pour un partisan de l'instrumentalisme interprétatif⁸, en attribuant une attitude propositionnelle douée d'une propriété sémantique, on peut prédire mais non expliquer causalement le comportement d'un dispositif physique. Pour un partisan de la théorie de l'erreur⁹, toute attribution d'attitude propositionnelle douée de propriété sémantique à un dispositif physique est fausse. Pour un partisan du non-factualisme¹⁰, une telle attribution est dénuée de condition de vérité¹¹.

Mais un matérialiste qui souscrit au réalisme intentionnel ne peut admettre la thèse de l'inefficacité causale des propriétés sémantiques des représentations mentales. Il doit leur reconnaître une efficacité causale dans le processus comportemental par lequel la représentation mentale concourt à la production d'un geste corporel. Le test de l'authenticité (ou de la réalité) d'une propriété, c'est son efficacité causale. Ce qui atteste la réalité d'une propriété, c'est sa capacité à produire des effets décelables : si la propriété sémantique d'une représentation mentale est une propriété authentique, alors elle doit concourir à produire des effets dans le processus comporte-

8. Voir Dennett, 1971, 1987.

9. Voir Churchland, 1981.

10. Voir Stich, 1983.

11. Pour une discussion détaillée des versions de l'irréalisme intentionnel, voir Jacob, 1997, ch. I.

mental. Dans la première partie de ce texte, je vais examiner un argument sur lequel s'appuient les adversaires matérialistes du réalisme intentionnel pour rejeter la thèse de l'efficacité causale des propriétés sémantiques des représentations mentales. Ayant indiqué la voie que, selon moi, doit suivre un matérialiste qui veut souscrire au réalisme intentionnel, j'examinerai dans la dernière partie le dilemme auquel se trouve confronté, selon Kim¹², le matérialiste qui souscrit au réalisme intentionnel.

2. La menace externaliste sur l'efficacité causale du contenu mental

On peut reconstruire le raisonnement qui conduit des philosophes matérialistes à adopter l'une des versions de l'irréalisme intentionnel de la manière suivante.

L'argument en faveur de l'irréalisme intentionnel :

1) Une explication causale doit faire référence à une propriété causalement efficace de la cause *c* dans le processus par lequel elle a produit son effet *e*.

2) Les seules propriétés causalement efficaces d'une cause *c* dans la production de son effet *e* sont des propriétés intrinsèques de *c* ou des propriétés de *c* qui dépendent systématiquement de propriétés sous-jacentes intrinsèques de *c*¹³.

3) Or, le contenu est une propriété historique et extrinsèque d'une représentation (mentale ou non mentale) qui ne dépend pas systématiquement de propriétés sous-jacentes intrinsèques de la représentation.

4) Donc : le contenu d'une représentation *R* d'un individu n'est pas une propriété causalement efficace dans le processus par lequel *R* contribue à produire causalement un comportement intentionnel de l'individu.

J'accepte les trois prémisses mais je rejette la conclusion (4). Je prétends que (4) n'est pas une conséquence des prémisses (1)-(3). Mais avant d'expliquer pourquoi je maintiens que (4) ne suit pas des prémisses (1)-(3), je vais d'abord motiver mon adhésion à la vérité des prémisses (1)-(3). Autrement dit, je vais préciser la nature de la menace externaliste sur la thèse de l'efficacité causale des propriétés sémantiques des représentations mentales.

12. Voir Kim, 1993a.

13. Par « dépendance systématique », j'entends ce que les anglophones nomment *supervenience* : une propriété *B* dépend systématiquement d'une propriété sous-jacente *A* si, nécessairement, chaque fois que *B* est exemplifiée, *B* est exemplifiée et (nécessairement) si *B* est exemplifiée, alors *A* est exemplifiée. Je mets la seconde occurrence de l'adverbe de modalité « nécessairement » entre parenthèses, car je ne veux pas trancher entre une interprétation faible ou forte de la notion de dépendance systématique requise ici. En qualifiant les propriétés physiques sous-jacentes d'« intrinsèques », je ne considère que les versions locales de la dépendance systématique et j'écarte l'interprétation globale. Voir Kim, 1987.

La vérité des prémisses (1) et (2) repose sur un principe que je tiens pour hautement plausible et qu'on pourrait appeler le « principe de l'inéquité de l'efficacité causale des propriétés d'une cause » et selon lequel l'efficacité causale n'est pas distribuée équitablement parmi les propriétés d'une cause. Pour corroborer ce principe, je gloserai un célèbre exemple dû à Dretske¹⁴.

a) Le sens du son

En chantant la Traviata, la Callas émet un son aigu *c* à l'instant *t*. L'émission du son à l'instant *t* brise le verre d'une vitre à l'instant *t* + 1. Appelons *c* l'émission du son à l'instant *t* et *e* le bris du verre à *t* + 1. *c* produit *e*. *c* exemplifie, d'une part, une propriété acoustique *K* et, d'autre part, un sens (ou un contenu) *F*. Posons la question suivante : *c* a-t-il produit *e* en vertu du fait qu'il exemplifiait la propriété *F*? Le fait que *c* était *F* est-il causalement efficace dans le processus par lequel *c* a produit *e*? Intuitivement, la réponse semble clairement négative. Si le son émis *c* avait eu un autre sens *F** (ou si *c* avait été dénué de sens), *c* aurait néanmoins pu produire *e*. La propriété causalement efficace de *c* dans le processus par lequel la vitre a été brisée (et *e* est advenu) est la propriété acoustique *K* : c'est parce que *c* exemplifiait la propriété acoustique *K* que *c* a produit *e*. Si *c* n'avait pas exemplifié *K*, *c* aurait pu ne pas produire *e* — le son aurait pu ne pas briser la vitre. La raison pour laquelle le fait que *c* était *K* a eu une efficacité causale dans le processus par lequel la vitre a été brisée est qu'il existe une corrélation nomique entre la fragilité du verre de la vitre et la propriété acoustique *K* du son et que la relation causale entre *c* et *e* peut être subsumée sous cette corrélation nomique. En revanche, il n'existe pas de corrélation nomique entre la fragilité du verre de la vitre et le fait que *c* exemplifie la propriété sémantique *F*. La propriété causalement efficace du son *c* est sa propriété nomique — la propriété grâce à laquelle l'interaction causale singulière entre le son et le verre de la vitre exemplifie une corrélation nomique.

Premièrement, cet exemple corrobore le principe de l'inéquité de l'efficacité causale des propriétés d'une cause présupposé par la première prémisses de l'argument en faveur de l'irréalisme intentionnel. Deuxièmement, cet exemple corrobore la vérité de la prémisse (2) de l'argument pour l'irréalisme intentionnel. Cette prémisse se sert du contraste entre les propriétés intrinsèques et les propriétés extrinsèques d'une entité *x*. Quoique ce contraste soit difficile à définir avec précision, il n'en est pas moins utile. Intuitivement, une propriété d'une entité *x* peut être qualifiée d'« intrinsèque » si son exemplification à l'instant *t* ne dépend de rien d'autre que de *x* à l'instant *t*. Une propriété est extrinsèque si elle n'est pas intrinsèque. En un sens, contrairement à la propriété acoustique *K* du son *c*, le sens (ou la propriété sémantique) *F* n'est pas une propriété intrinsèque de *c* : en effet, le son

14. Voir Dretske, 1988, p. 79.

c n'aurait pas été le même son s'il n'exemplifiait pas la propriété *K*. Mais le son *c* pourrait demeurer le même son s'il n'avait pas la propriété sémantique *F*. Certes, on pourrait faire valoir que l'événement consistant à produire le son *c* ne serait pas le même événement si la propriété sémantique *F* faisait défaut au son *c*. Mais il y a un sens où deux émissions sonores pourraient être deux exemplaires (ou deux occurrences) du même son-type même si l'une des deux émissions seulement exemplifiait la propriété sémantique *F*. En revanche, deux émissions sonores ne pourraient pas être deux exemplaires du même son-type à moins qu'elles n'exemplifient toutes les deux la même propriété acoustique *K*. Un son qui aurait la même propriété sémantique *F* et exemplifierait une propriété acoustique *K** distincte de *K* ne serait pas le même son. Contrairement à sa propriété acoustique *K*, le son tient sa propriété sémantique *F* des intentions du compositeur, Giuseppe Verdi. Certes, le son a été choisi par Verdi. Mais contrairement au fait que le son exemplifie le sens *F*, le fait que le son exemplifie la propriété acoustique *K* est indépendant des intentions de Verdi. Verdi a donné au son son sens. Il n'a pas créé la gamme de toutes pièces. La propriété sémantique *F* du son est donc une propriété historique extrinsèque du son.

Selon la distinction entre l'intentionnalité primitive et l'intentionnalité dérivée (à laquelle souscrivent les partisans du réalisme intentionnel), il existe évidemment une différence entre le contenu (linguistique) d'un son émis par une chanteuse et le contenu d'une représentation mentale (ou d'un état cérébral) d'un être humain. Le second est primitif. Le premier est dérivé du second¹⁵. D'une manière générale, la signification d'une représentation non mentale dépend du contenu d'une représentation mentale. Mais le contenu mental lui aussi est une propriété historique et extrinsèque d'un état cérébral de l'individu. Il dépend de certaines corrélations entre le cerveau de l'individu et les propriétés exemplifiées dans son environnement¹⁶. Ce qui nous conduit tout droit à la prémisse (3) de l'argument en faveur de l'irréalisme intentionnel. Pour corroborer la prémisse (3), j'emprunterai un exemple à Dennett, que j'exploiterai dans un style dretskeén¹⁷.

b) Le distributeur automatique de boissons

Soit un distributeur automatique de boissons *S* situé en France. Il émet une boisson non alcoolisée chaque fois qu'il reçoit une pièce de cinq francs. Appelons *m* l'émission d'une boisson : *S* produit *m* grâce à l'ouverture d'un clapet. Pour simplifier, je traiterai la pièce de cinq francs *c* insérée dans la fente du dispositif *S* comme la cause de *m*, l'émission d'une boisson.

15. Pour une critique de cette distinction, voir Dennett, 1987, ch. 8. Pour une critique des critiques de Dennett, voir Jacob, 1997, ch. I.

16. Je ne fais ici qu'ébruiter mon adhésion à une composante informationnelle dans la conception du contenu mental.

17. Voir Dennett, 1987, ch. 8 et Dretske, 1998.

Aussi, c a des propriétés physiques intrinsèques P (une masse, une forme géométrique, une composition physique caractéristique, etc.). Et c a aussi une valeur monétaire V . V est manifestement une propriété historique extrinsèque de c parce que ce qui confère V à c , c'est le fait historique d'avoir été estampillée par une personne travaillant pour la Banque de France, et non par une autre personne — en particulier, non par un faussaire. Ce qui confère V à c , c'est le fait d'être le relatum d'une relation historique appropriée avec un agent de la Banque de France. Supposons qu'un disque métallique c^* soit physiquement et chimiquement indiscernable de c . Si, contrairement à c , c^* n'a pas bénéficié d'une relation historique et extrinsèque avec un agent de la Banque de France, alors en dépit du fait que c^* exemplifie les propriétés physiques intrinsèques P , V fait pourtant défaut à c^* . Dire de V que c'est une propriété historique et extrinsèque de c , c'est donc dire que le fait que c est V ne dépend pas systématiquement du fait que c exemplifie la propriété physique intrinsèque sous-jacente P^{18} .

L'insertion de c à l'instant t a incontestablement produit m (l'émission d'une boisson) à l'instant $t + 1$. L'insertion d'un objet physique qui n'aurait pas les propriétés physiques intrinsèques d'une pièce de cinq francs dans la fente de S ne provoquerait pas m : S n'émettrait pas de boisson. Mais posons à présent la question suivante : la propriété V de c est-elle causalement efficace dans le processus par lequel S produit m ? Le fait que c est V est-il causalement responsable du fait qu'un clapet dans S libère une boisson?

Intuitivement, la réponse est : Non. Le fait que c est V n'est pas responsable du fait que S produit m . Si c contribue à produire m , c'est parce que c est P et non pas parce que c est V . Ce n'est pas parce que S « croit » avoir reçu une pièce de cinq francs qu'il émet une boisson ; c'est parce qu'il détecte un disque métallique ayant la propriété physique intrinsèque P . L'ouverture du clapet dans S est reliée nomologiquement à la propriété P de c , non à sa propriété V . Le distributeur automatique de boissons ne détecte pas la valeur monétaire V ; il réagit aux propriétés physiques des objets insérés dans sa fente. La propriété V de c est donc « épiphénoménale » dans le mécanisme de production de m par c . Si la valeur monétaire V du disque métallique inséré dans la fente du distributeur est un modèle du contenu mental, alors ce que je crois quand je crois que le verre placé devant moi contient de l'eau n'explique pas du tout ce que je fais lorsque je transporte ma main droite au voisinage du verre. Les prémisses (2)-(3) de l'argument en faveur de l'irréalisme intentionnel semblent donc acceptables. La conclusion (4) n'est-elle pas inévitable? Comment puis-je accepter les prémisses et rejeter la conclusion?

18. Du moins, c'est dire que V ne dépend pas systématiquement de P dans la version forte de la dépendance systématique, c'est-à-dire qu'il n'est pas vrai que dans tous les mondes possibles, si x est P , alors x est V .

Le distributeur automatique de boissons est un dispositif physique qui admet en « entrée » des disques métalliques ayant la propriété physique intrinsèque *P* et la propriété extrinsèque *V* et répond en émettant une boisson. La question que nous nous sommes posée était la suivante : le fait que *c* est *V* est-il causalement responsable dans le mécanisme par lequel un clapet s'ouvre dans *S* lorsque *c* est insérée dans la fente de *S* et *S* émet une boisson? La propriété *V* de *c* est-elle causalement efficace dans le mécanisme grâce auquel *S* produit *m* (l'émission d'une boisson)? Nous venons de conclure que la réponse à cette question est négative. Appelons cette question la question 1. Je maintiens que la réponse à la question 1 n'a aucune raison de mentionner la propriété *V* de *c*. Autrement dit, la propriété de *c* qui explique la production du mouvement physique *m* est la propriété *P*.

Or, cette question porte sur le mécanisme de la production du mouvement physique, *m*¹⁹. Comprendre le mécanisme de production du mouvement physique *m*, n'est-ce pas expliquer le comportement du distributeur de boissons? Suivant les indications de Dretske²⁰, je voudrais suggérer une réponse négative à cette question en considérant une question différente de la question 1. Appelons cette question la question 2 : pourquoi chaque fois qu'un disque métallique ayant la propriété *P* est inséré dans la fente de *S*, *S* produit-il *m*, l'émission d'une boisson? Pourquoi l'insertion d'un disque métallique ayant la propriété *P* dans une fente de *S* produit-elle régulièrement *m*, la chute d'une boisson? En un mot, pourquoi le comportement de *S* est-il constitué par (ou composé de) la coordination entre *c* et *m*?

En adoptant la conception « componentielle » du comportement défendue par Dretske²¹, disons que le comportement de *S* est le processus par lequel (ou au cours duquel) l'insertion d'un disque métallique *c* produit l'émission d'une boisson, *m*. Selon cette conception, ce serait une confusion conceptuelle que de confondre la « sortie » *m* avec le comportement. L'émission d'une boisson, *m*, est l'un des constituants du comportement de *S*. Ce n'est pas le comportement lui-même²². L'insertion du disque métallique *c* est l'autre constituant du comportement de *S*. Ces deux constituants sont unis par la relation causale : l'insertion de *c* à l'instant *t* est la cause de *m*, l'émission d'une boisson à l'instant *t* + 1. L'insertion de *c* à l'instant *t* n'est donc pas, contrairement à ce qu'affirme la conclusion (4) de l'argument en faveur de l'irréalisme intentionnel, une cause du comportement de *S*. C'est un constituant du comportement de *S*. C'est aussi une cause du second constituant du comportement de *S* : *m*, l'émission d'une boisson. Voilà pourquoi je peux

19. L'émission d'une boisson, *m*, est, pour le distributeur, ce qu'est le geste corporel pour un être humain.

20. Voir Dretske, 1988, ch. 1-2.

21. Voir Dretske, 1988, ch. 1-2, et Jacob, 1997, ch. VIII.

22. Contrairement à la conception du comportement qu'on trouve dans le fonctionnalisme et dans le monisme anomal, dans la conception componentielle, le geste corporel n'est pas identique au comportement.

admettre la vérité des trois prémisses de l'argument en faveur de l'irréalisme intentionnel sans accepter sa conclusion (4). Appelons (4*) la conclusion qui est une conséquence valide des prémisses (1)-(3) de l'argument en faveur de l'irréalisme intentionnel. La conclusion (4*) est la suivante :

(4*) Le contenu d'une représentation d'un individu n'est pas une propriété causalement efficace dans le processus par lequel une représentation contribue à produire un mouvement corporel de l'individu.

Ce que démontre l'exemple du distributeur automatique de boissons, c'est, comme l'affirme la conclusion (4*), que la propriété extrinsèque *V* du disque métallique *c* n'est pas causalement efficace dans la production de *m*, l'émission d'une boisson. La valeur économique « authentique » de la pièce de cinq francs n'est pas causalement efficace dans le processus d'ouverture du clapet dans *S*. Seule l'exemplification de la propriété physique intrinsèque *P* par le disque métallique *c* est causalement efficace dans ce mécanisme.

Comme j'ai soutenu que le comportement est un processus qui ne s'identifie à aucun de ses constituants et comme le mouvement physique (ou le geste corporel) *m* est l'un de ces constituants, je vais à présent faire valoir qu'il existe un sens dans lequel l'explication causale du comportement n'est pas la même chose que l'explication de la production du mouvement physique *m*. En réponse à la question 1 sur le mécanisme, j'ai fait valoir que seule la propriété physique intrinsèque *P* de *c* était causalement efficace pour assurer l'ouverture du clapet dans *S*. Mais la question 2 est distincte de la question 1. En posant la question 2, on demande : pourquoi chaque insertion d'un disque métallique *c* dans la fente de *S* provoque-t-elle régulièrement *m*, la chute d'une boisson (grâce à l'ouverture du clapet)? Pourquoi le distributeur automatique est-il un dispositif physique dont le comportement est constitué par la coordination entre deux événements : l'insertion de *c* et la chute d'une boisson?

Contrairement à la question 1, la question 2 est une question historique : poser la question 2, c'est se demander quel processus a donné naissance à la coordination entre *c* et *m* dans *S*. La réponse à la question 2 aura la forme générale suivante : comme *S* est un artefact fabriqué par un ingénieur humain, la coordination entre *c* et *m* a été conçue et réalisée par un agent intentionnel lui-même doué d'attitudes propositionnelles²³. De surcroît, je veux à présent suggérer qu'il existe un sens dans lequel le fait que le disque métallique *c* possède une valeur monétaire (la propriété extrinsèque *V*) contribue à répondre à la question 2, c'est-à-dire à expliquer la structure du comportement de *S*. En un mot, le fait que *c* est *V* contribue à expliquer la coordination entre *c* et *m* dans *S*.

23. Les activités par lesquelles un ingénieur a assuré la coordination entre *c* et *m* sont ce que Dretske, 1988, nomme les « causes structurantes » du comportement du distributeur.

Supposons en effet que, grâce aux efforts des agents de la Banque de France, il existe une corrélation fiable entre les exemplifications de la valeur monétaire V et les exemplifications des propriétés physiques par ces disques métalliques que sont les pièces de cinq francs²⁴. Compte tenu, d'une part, du prix des boissons et, d'autre part, de la corrélation fiable entre la valeur monétaire V et les propriétés physiques intrinsèques P de certains objets métalliques, les objets métalliques ayant la propriété P ont été choisis, recrutés ou sélectionnés par un ingénieur humain pour être régulièrement les causes de m parce que ces objets métalliques exemplifient V . C'est bien parce que les disques métalliques sont P qu'ils provoquent l'ouverture du clapet dans S . Mais c'est parce qu'un objet qui est P est V qu'il a été choisi par un ingénieur humain comme cause de l'ouverture du clapet dans S , c'est-à-dire comme cause de m . Sans doute la corrélation entre P et V n'est-elle pas parfaite : à l'occasion, l'insertion d'une « fausse » pièce de cinq francs — un disque métallique qui est P mais non V — provoque m , l'émission d'une boisson. Mais vraisemblablement, le bénéfice qu'entraînerait une amélioration de la sensibilité du distributeur ne compense pas le coût d'un tel effort.

Supposons à présent qu'à la date t , la Banque de France décide de changer les propriétés intrinsèques des pièces de cinq francs. Désormais, il n'y a plus de corrélation fiable entre les exemplifications de la propriété P et les exemplifications de la propriété V : après t , ce ne sont plus les disques métalliques ayant la propriété P qui exemplifient la propriété V ; les disques métalliques ayant la propriété V ont dorénavant la propriété physique intrinsèque P^* . Dans cette hypothèse et en supposant que le prix des boissons non alcoolisées reste constant, l'insertion d'un disque métallique c ayant la propriété intrinsèque P dans la fente de S ne contribue plus à la production de m . Avant la décision de la Banque de France, l'ingénieur humain avait sélectionné des objets métalliques c ayant la propriété P comme cause de m . Après la décision de la Banque de France — en admettant la fixité du prix des boissons —, les ingénieurs français doivent fabriquer de nouveaux distributeurs automatiques : ils sélectionnent désormais des objets métalliques c^* ayant la propriété P^* (différente de P) comme cause de m .

Dans cette hypothèse, on peut donc dire que de nouveaux disques métalliques c^* ayant certaines propriétés physiques intrinsèques P^* ont été choisies pour être introduites dans la fente d'un nouveau distributeur S^* parce que ces disques exemplifient la propriété extrinsèque V . Par parité, c'est parce que les disques c ayant la propriété P exemplifiaient la propriété V qu'ils avaient été choisis initialement comme cause de m dans S . C'est donc parce que certains objets métalliques ont la propriété V qu'ils ont été sélectionnés dans le processus de production de m . Si les objets métalliques c ayant la propriété intrinsèque P n'exemplifiaient pas en général

24. La corrélation fiable entre les exemplifications de P et de V correspond à l'interprétation faible de la dépendance systématique entre V et P (au sens de Kim).

la propriété V , ces objets n'auraient pas été sélectionnés comme cause de m . Le fait qu'un disque métallique exemplifie la propriété extrinsèque V contribue à expliquer la coordination entre c et m dans le comportement de S . L'exemplification de la propriété extrinsèque V contribue donc à expliquer la structure du comportement de S .

c) La proie

Dans le cas du distributeur automatique de boissons, l'explication historique de la structure du comportement (ou de la coordination entre c et m) fait référence à un processus intentionnel de sélection : c'est un agent humain doué de désirs et de croyances qui a assuré cette coordination. Je veux à présent considérer brièvement le cas hypothétique dans lequel une représentation sensorielle d'un animal serait recrutée comme cause de mouvements corporels par un processus non intentionnel de sélection : la sélection naturelle à l'œuvre dans l'évolution des espèces. Soit une proie B (pour biche). B exécute des mouvements de fuite m à l'instant t . Je supposerai que le comportement de fuite de la proie B est un comportement intentionnel non volontaire. Nous pouvons poser au moins trois questions distinctes au sujet des mouvements de fuite m de B .

Premièrement, nous pouvons demander : pourquoi B accomplit-elle m à l'instant t ? À cette question, on répondra : parce qu'à l'instant $t - 1$, B a perçu le prédateur P dans le voisinage. Appelons c l'état interne grâce auquel B a détecté P . La présence du prédateur P est la cause de l'occurrence de l'état interne c de B . Comme l'état c de B est lui-même la cause du mouvement de fuite m , Dretske qualifie la présence du prédateur de « cause déclenchante » du comportement de fuite de B ²⁵.

Deuxièmement, nous pouvons poser la question non historique suivante : quel est le mécanisme interne grâce auquel l'occurrence de l'état interne c dans B produit les mouvements corporels m ? C'est l'équivalent de la question 1 dans le cas du comportement du distributeur automatique de boissons. À cette question, la réponse pertinente sera la réponse suivante : les neurones moteurs dans le cortex moteur de B sont reliés aux neurones sensoriels du système visuel (ou olfactif) de B . Lorsque les neurones sensoriels déchargent, ils envoient un signal aux neurones moteurs. La décharge de ces derniers gouverne la contraction des muscles des membres inférieurs de B .

Enfin, nous pouvons poser la question historique suivante : comment se fait-il que chaque fois que B est dans l'état c , B produit m ? Pourquoi l'état c de B a-t-il été recruté comme cause de m , des mouvements de fuite, plutôt que comme cause d'autre chose? Lorsque B est dans l'état c , B s'enfuit. Pourquoi B ne combat-elle pas son prédateur? Ou pourquoi ne reste-t-elle pas immobile? La réponse à cette question historique doit venir en

25. Voir Dretske, 1988.

deux étapes complémentaires. Premièrement, *c* a été recruté comme cause de *m* (des mouvements de fuite) par la sélection naturelle qui, comme je l'ai dit, est un processus sélectif non intentionnel. Dans la compétition pour la survie entre des ancêtres de la proie *B* et d'autres animaux qui étaient aussi les proies des ancêtres du prédateur *P*, la sélection naturelle a favorisé les ancêtres de la proie *B* qui avaient une coordination entre *c* et *m*. Les proies des ancêtres de *P* qui étaient dépourvues de cette coordination entre *c* et *m* n'ont pas survécu à la sélection et furent dépourvus de descendance²⁶. Deuxièmement, la sélection naturelle n'aurait pas recruté *c* comme cause de *m* chez les ancêtres de *B* si *c* n'avait pas été un indicateur fiable de la présence du prédateur *P*. La propriété d'être un indicateur fiable n'est sans doute pas une propriété pleinement sémantique ; c'est une propriété proto-sémantique. En tout cas, c'est une propriété historique et extrinsèque d'un état interne de la proie qui suppose l'existence de corrélations nomiques entre des états du cerveau de la proie et l'exemplification de la propriété d'être un prédateur dans son environnement ou d'une propriété coextensionnelle avec cette propriété²⁷.

Un matérialiste peut donc souscrire au réalisme intentionnel et tenir la propriété sémantique d'une représentation mentale pour causalement efficace dans le processus comportemental par lequel la représentation contribue à produire un geste corporel. Il le peut, à condition d'adopter conjointement la conception componentielle du comportement et ce que, suivant Kim, j'ai appelé ailleurs « la stratégie de la dualité des *explananda*²⁸ », c'est-à-dire à condition de distinguer les questions 1 et 2 : la question du mécanisme de la production d'un mouvement physique et la question de la sélection de la coordination grâce à laquelle un état interne est recruté comme cause du mouvement physique²⁹.

26. Suivant Sober, 1984, Dretske, 1988, ch. 4, admet une distinction tranchée entre l'explication causale du comportement intentionnel d'un individu et une explication sélectionniste de la distribution d'un trait dans une population. Dans Jacob, 1997, ch. VIII, section 6, et dans Jacob, 1998, j'ai longuement expliqué pourquoi, à la suite de Neander, 1988, 1995, je pense qu'une explication historique de la coordination entre un état interne *c* et un mouvement physique *m* est conjointement une explication causale et une explication sélectionniste.

27. Pour le partisan que je suis d'une conception téléosémantique à base informationnelle du contenu mental — contenu conceptuel et contenu non conceptuel —, le fait qu'un état interne soit un indicateur fiable est une condition nécessaire mais non suffisante du fait que cet état est une représentation mentale douée d'une propriété sémantique authentique. Voir Jacob, 1997, ch. IV et VIII.

28. Voir Kim, 1990, 1991, et Jacob, 1997, ch. VIII.

29. Comme me l'a fait remarquer à juste titre un lecteur de la revue, je n'ai considéré que le rôle causal du contenu d'une représentation sensorielle innée, et non le contenu d'une représentation conceptuelle acquise. Comme je l'indique à la note 5, à la différence de Dretske, je soutiens que le contenu d'une représentation innée peut contribuer à expliquer un comportement intentionnel non volontaire. Pour la contribution du contenu d'une représentation conceptuelle acquise au comportement intentionnel volontaire, voir Jacob, 1997, ch. VIII, et Jacob, 1998.

3. Le réalisme intentionnel face à la menace de l'exclusion explicative

Comme je l'ai soutenu ailleurs, un matérialiste qui souscrit au réalisme intentionnel est confronté à ce que j'ai appelé « le dilemme du réalisme intentionnel³⁰ ». Ce dilemme surgit de la tension entre deux desiderata : une tendance réductionniste et une tendance antiréductionniste.

D'une part, comme nous venons de le voir dans la section précédente, un partisan du réalisme intentionnel se sent tenu de conférer une efficacité causale originale aux propriétés sémantiques des états cérébraux d'un individu. Son réalisme le pousse à accorder aux faits sémantiques une responsabilité causale particulière dans la production de faits physiques : un système physique capable de former des représentations douées de propriétés sémantiques doit pouvoir produire des effets physiques originaux. D'autre part, un matérialiste ne peut admettre que les faits sémantiques soient des faits métaphysiques bruts ou ultimes. Le fait qu'un état cérébral possède un contenu (ou une propriété sémantique) ne peut pas être une donnée métaphysique irréductible. C'est ce qu'exprime Fodor lorsqu'il dit qu'un réaliste intentionnel ne peut que souscrire au réductionnisme et que les propriétés sémantiques n'ont aucune chance de figurer dans le « catalogue ultime » des propriétés répertoriées par la physique fondamentale³¹. C'est aussi ce qu'exprime Dennett lorsqu'il compare toute explication qui fait référence au contenu à un emprunt : quiconque fait un emprunt sémantique est tenu de le rembourser en termes non sémantiques³². C'est aussi ce que dit Field qui tient la doctrine qu'il nomme le « sémanticalisme » pour incompatible avec le matérialisme³³. Un matérialiste partisan du réalisme intentionnel est donc tiraillé entre deux projets : il cherche à résoudre le problème de « la causalité mentale » et il cherche à combler le fossé entre les propriétés sémantiques et les propriétés non sémantiques d'un dispositif physique (c'est le projet de la « naturalisation de l'intentionnalité »).

Dans une série d'écrits importants³⁴, Kim a mis le matérialiste qui est aussi un partisan du réalisme intentionnel au défi de répondre à l'argument de « l'exclusion explicative ». Cet argument est destiné à convaincre le matérialiste de renoncer à la composante antiréductionniste du réalisme intentionnel. L'argument de l'exclusion explicative procède en deux étapes dont la première consiste à montrer au partisan de ce que Kim nomme le « matérialisme non réductionniste » — par exemple, le partisan du physicalisme des exemplaires, par opposition au physicalisme des types — que toute chaîne causale « horizontale » impliquant l'exemplification de deux propriétés

30. Voir Jacob, 1997, ch. I.

31. Voir Fodor, 1987, p. 97.

32. Voir Dennett, 1971.

33. Voir Field, 1972.

34. Dont bon nombre sont recueillis dans Kim, 1993b. Voir aussi Kim, 1998.

mentales (ou sémantiques) suppose que l'exemplification de la propriété mentale qui joue le rôle de cause est inévitablement impliquée dans une chaîne causale « descendante³⁵ ». Un matérialiste non réductionniste soutient que la propriété mentale d'un état cérébral (par exemple, la propriété sémantique ou le contenu) n'est pas identique ou réductible à une propriété physique intrinsèque de l'état cérébral³⁶. Un matérialiste non réductionniste peut donc souscrire au physicalisme des exemplaires et identifier un exemplaire d'une représentation mentale (d'une attitude propositionnelle, par exemple) à un état cérébral concret de l'individu.

L'exemplification d'une propriété mentale (ou sémantique) *M* peut être un maillon dans une chaîne causale « horizontale » lorsqu'elle provoque l'exemplification d'une autre propriété mentale (ou sémantique) *M**. Lorsqu'une pensée produit une autre pensée, on est en présence d'une chaîne causale « horizontale » de ce genre. L'exemplification d'une propriété mentale (ou sémantique) *m* peut être un maillon dans une chaîne causale « descendante » lorsqu'elle provoque l'exemplification d'une propriété physique *P**. Lorsqu'une représentation mentale produit un geste corporel, on est en présence d'une chaîne causale « descendante » de ce genre³⁷.

Supposons donc que l'exemplification de la propriété sémantique *M* provoque l'exemplification d'une autre propriété sémantique *M**. Dans ce qu'on pourrait appeler « le carré de Kim », il est évidemment crucial de supposer que les propriétés sémantiques (ou mentales) sont distinctes des propriétés physiques sous-jacentes dont elles dépendent. Un partisan du matérialisme non réductionniste supposera naturellement que la propriété *M** dépend systématiquement d'une propriété physique non sémantique sous-jacente *P** ou — ce qui revient au même — que *P** sert de base physique à la réalisation de *M**. Selon les hypothèses du matérialiste non réductionniste, nous nous trouvons en présence d'un cas illustré par la Figure 1.

À cette étape de l'argument, Kim fait valoir que l'exemplification de *M** semble donc provoquée conjointement par l'exemplification de la propriété de *M* et par l'exemplification de *P** : la propriété sémantique *M** est exemplifiée à la fois parce qu'une autre propriété sémantique *M* a été exemplifiée et parce que sa base physique *P** est exemplifiée. Kim maintient qu'on ne peut ni envisager que l'exemplification de *M* et l'exemplification de *P** collaborent causalement à l'exemplification de *M** en faisant valoir que, du

35. Voir Kim, 1993a.

36. Un psychofonctionnaliste ou un partisan du monisme anomal est un représentant du matérialisme non réductionniste.

37. Par souci de complétude, on peut mentionner l'existence hypothétique de chaînes causales « ascendantes » dans lesquelles peuvent entrer des propriétés sémantiques : tels seraient les cas où l'exemplification de *M* provoquerait l'exemplification d'une propriété « émergente » sociale. Voir Kim, 1993a, p. 351. Ces cas ne joueront aucun rôle dans la suite.

Figure 1

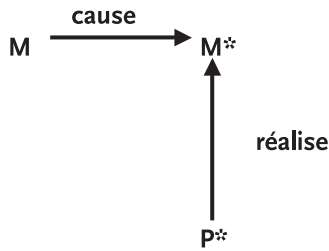
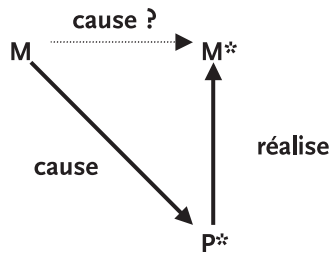


Figure 2



point de vue du matérialisme non réductionniste, l'exemplification de P^* à elle seule est une condition suffisante de l'exemplification de M^* ³⁸. Pour la même raison, Kim rejette l'interprétation selon laquelle l'exemplification de M^* serait « surdéterminée » par deux facteurs causaux. Il conclut la première étape de l'argument en indiquant que la seule voie ouverte au matérialiste non réductionniste est de supposer que l'exemplification de M contribue à produire l'exemplification de M^* grâce au fait qu'elle provoque l'exemplification de P^* , comme dans la Figure 2. Cette figure montre que, du point de vue du matérialisme non réductionniste, la causalité « horizontale » entre une exemplification de M et une exemplification de M^* dépend de (ou suppose) la causalité « descendante » entre une exemplification de M et une exemplification de P^* , la base physique de M^* .

La seconde étape de l'argument de l'exclusion explicative consiste à faire valoir que M elle-même doit aussi dépendre systématiquement d'une propriété physique non sémantique sous-jacente P , comme l'indique la Figure 3.

Dans la production de l'exemplification de P^* , l'exemplification de M entre donc en compétition avec l'exemplification de sa propre base physique P . Selon le « principe de la clôture causale du domaine physique »,

38. Si on suppose que M^* dépend de P^* dans l'interprétation forte de la dépendance systématique, cela va de soi puisque, dans cette interprétation, il est nécessaire (ou vrai dans tous les mondes) que si P^* est exemplifiée, alors M^* est exemplifiée aussi. Voir Kim, 1984.

Figure 3

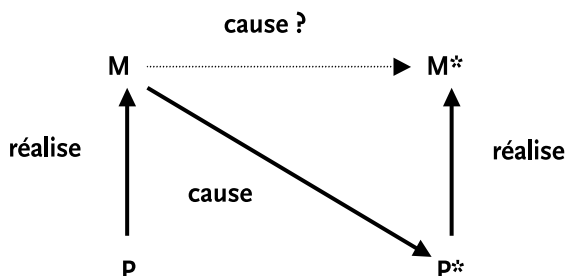
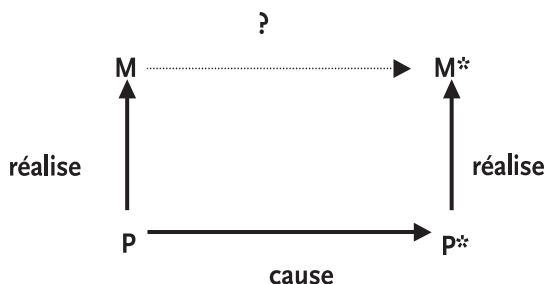


Figure 4



tout ancêtre causal ou tout héritier causal d'une exemplification d'une propriété physique doit être lui-même une exemplification de propriété physique. Dans la compétition pour produire l'exemplification de P^* , l'exemplification de M semble inévitablement devoir être privée d'efficacité causale par l'efficacité de l'exemplification de P . On se retrouve dans la situation illustrée par la Figure 4 dans laquelle la propriété P réalise la propriété M (ou réciproquement M dépend systématiquement de P), P^* réalise M^* et l'exemplification de P est la cause de l'exemplification de P^* . L'exemplification de M semble dépourvue de tout rôle causal. Comme le dit Kim³⁹, « dans le cas de la relation causale alléguée entre M et M^* , la situation est comparable à une série d'ombres projetées par une voiture en mouvement : il n'y a pas de relation causale entre l'ombre de la voiture à l'instant t et son ombre à l'instant suivant ; l'une et l'autre sont l'effet du mouvement de la voiture. Le mouvement de la voiture est un processus causal authentique, mais la séquence des ombres projetées, aussi régulière et nomologique qu'elle soit, n'est pas un processus causal⁴⁰ ».

39. Kim, 1998, p. 45.

40. La distinction entre processus causal authentique et pseudo-processus vient de Salmon, 1984. Voir Jacob, 1997, ch. VII.

S'il veut éviter la conclusion de Kim, le partisan du matérialisme non réductionniste peut sans doute résister à la première étape de l'argument qui vise à démontrer que la causalité « horizontale » entre l'exemplification de M et l'exemplification de M^* implique inévitablement la causalité « descendante » entre l'exemplification de M et l'exemplification de P^* , la base physique de M^* . Du moins peut-il rejeter le modèle de la causalité « descendante » que lui prête Kim. Kim traite en effet la contribution de P^* à l'exemplification de M^* comme un cas de relation causale en tous points comparable à la contribution supposée de M^{41} . Mais comme Kim lui-même l'a fait remarquer à juste titre à Searle⁴², c'est manifestement une erreur de traiter la relation de dépendance systématique entre la propriété M^* et sa base physique sous-jacente P^* comme une relation causale : même si l'exemplification de P^* est une condition suffisante de l'exemplification de M^* , la première n'est pas la cause de la seconde⁴³. Premièrement, dans une relation causale, la cause doit normalement précéder son effet. Or, l'exemplification de M^* et l'exemplification de sa base physique P^* sont simultanées. De surcroît, si on admet la version forte de la dépendance systématique entre l'exemplification de M^* et celle de sa base physique P^* , non seulement les deux exemplifications sont simultanées, mais la relation entre les deux exemplifications ne peut pas, comme une relation causale singulière, être une relation contingente : tout ce qui est P^* à l'instant t est nécessairement M^* à cet instant. Dans tous les mondes possibles, si quelque chose exemplifie P^* à l'instant t , cette chose doit exemplifier M^* à cet instant. Les deux exemplifications sont reliées de manière nomologique, mais la loi qui les relie n'est pas une loi causale. Cette loi ne peut pas être une loi causale en raison de la simultanéité entre l'exemplification respective des deux propriétés impliquées. C'est une identité nomique du type « eau = H₂O ».

L'exemplification de M est dans une relation causale avec l'exemplification de M^* . Quoique l'exemplification de P^* soit une condition suffisante de l'exemplification de M^* , la première n'est pas une cause de la seconde. L'exemplification de M et l'exemplification de P^* font donc une contribution différente à l'exemplification de M^* . L'efficacité causale de l'exemplification de M dans la production causale de l'exemplification de M^* n'est donc pas menacée par l'exemplification de P^* . Le partisan du matérialisme non réductionniste peut rejeter la thèse de Kim selon laquelle toute relation causale « horizontale » entre M et M^* implique une relation causale « descendante » entre M et P^* , la base physique de M^* . Ce que le partisan du matérialisme non

41. Dans la note 30, p. 351, Kim, 1993a, écrit de la rivalité entre les rôles respectifs de P^* et de M à l'exemplification de M^* « qu'elle est essentiellement identique à la situation dans laquelle nous sommes lorsque nous sommes en présence de deux causes distinctes d'un seul et même événement, chacune pouvant prétendre être une cause suffisante ».

42. Voir Kim, 1998, p. 47-50. La critique de Kim est dirigée contre Searle, 1992.

43. Slors, 1998, fait la même observation.

réductionniste objecte au modèle kimien de la causalité « descendante », c'est la supposition selon laquelle la chaîne causale reliant l'exemplification de M à l'exemplification de M^* passe par deux relations causales intermédiaires : premièrement, la relation causale entre l'exemplification de M et l'exemplification de P^* et deuxièmement, la relation causale entre l'exemplification de P^* et l'exemplification de M^* . Le partisan du matérialisme non réductionniste peut, d'une part, soutenir que produire l'exemplification de M^* , ce n'est rien d'autre que produire l'exemplification de P^* , puisque les deux propriétés sont co-exemplifiées simultanément. En produisant l'une, l'exemplification de M produit ipso facto l'autre. Il peut, d'autre part, soutenir que l'exemplification de P^* est nomologiquement constitutive de l'exemplification de M^* : M^* est exemplifiée en vertu du fait que P^* est exemplifiée ; M^* ne serait pas exemplifiée si P^* n'était pas exemplifiée au sens où un liquide ne serait pas de l'eau s'il n'était pas composé de molécules de H_2O . Mais l'exemplification de P^* n'est pas la cause de l'exemplification de M^* . L'exemplification de M peut donc rester la cause « horizontale » de l'exemplification de M^* sans que son efficacité causale soit menacée par celle de P^* .

Ce que répudie le partisan du matérialisme non réductionniste dans le modèle de la causalité « descendante » que lui impute Kim, c'est l'hypothèse d'une sous-relation causale « ascendante » entre l'exemplification de M^* et l'exemplification de sa base physique, P^* . Mais, même si le partisan du matérialisme non réductionniste peut désarmer la première étape de l'argument de l'exclusion explicative et, non sans bonnes raisons, refuser d'admettre que la causalité mentale « horizontale » implique le modèle de la causalité « descendante » entre une exemplification de propriété sémantique et une exemplification de propriété physique sous-jacente que lui impute Kim, il n'est pourtant pas au bout de sa peine⁴⁴. En effet, comme l'indique la seconde étape de l'argument, lorsque l'exemplification d'une propriété sémantique contribue à produire l'exemplification d'une propriété physique, la seule application du principe de la clôture causale du domaine physique suffit à faire surgir la menace de l'exclusion explicative.

Comme l'a montré la section précédente, la conception componentielle du comportement et la stratégie de la dualité des *explananda* semblent permettre au matérialiste qui souscrit au réalisme intentionnel de parer à la menace externaliste et de conférer une efficacité causale aux propriétés sémantiques dans l'explication de la structure du comportement, c'est-à-dire de la coordination entre un état interne et un mouvement physique. Mais la stratégie de la dualité des *explananda* est-elle vraiment compatible avec le matérialisme? Cette stratégie s'appuie sur un contraste entre deux sortes de faits à expliquer : le mouvement physique et le comportement. L'explication du

44. Devant l'acte d'accusation de Kim, j'ai voulu défendre le dossier du matérialiste anti-réductionniste aussi loin que possible. Mais, comme l'indique la suite du texte, je reconnais la validité de la seconde étape de l'argument de Kim.

mécanisme de la production du mouvement physique (la réponse à ce que j'ai appelé la question 1) fait référence aux propriétés non sémantiques de l'état interne *c* — aux propriétés physiques, chimiques et neurophysiologiques intrinsèques de l'état *c* (dans le cas où *c* est l'état interne d'un animal). L'explication de la structure du comportement ou de la coordination entre *c* et *m* (la réponse à la question 2) fait référence à une propriété sémantique ou proto-sémantique de *c*. Mais comme l'a fait valoir Kim⁴⁵, le matérialiste partisan du réalisme intentionnel ne peut pas se contenter de faire valoir que les propriétés sémantiques d'un état interne peuvent contribuer à expliquer l'un des deux *explananda*. Le fait qu'il existe une « explication par des raisons » pour la structure du comportement ne suffit pas à écarter la menace de l'exclusion explicative par une explication physique ou neurophysiologique⁴⁶. Il doit de surcroît montrer pour chacun des deux *explananda* qu'il n'est pas susceptible de recevoir deux explications : une « explication par des raisons » et une « explication physique ». La seule voie qui paraît ouverte au partisan de la stratégie de la dualité des *explananda* est de faire valoir que la structure du comportement au sens de la conception componentielle ne peut pas recevoir d'explication physique ou neurophysiologique.

Celui qui soutient que la structure du comportement au sens de la conception componentielle n'est pas susceptible de recevoir une explication physique ou neurophysiologique semble confronté au dilemme suivant : soit il suppose que les faits et les phénomènes qu'expliquent les propriétés sémantiques — à savoir la structure du comportement — ne sont pas des faits et des phénomènes physiques ou neurophysiologiques mais sont des faits non physiques. Auquel cas, il semble condamné à répudier purement et simplement le monisme matérialiste et à admettre une forme de dualisme des substances de type cartésien. Soit il reconnaît que les faits constitutifs du comportement au sens de la conception componentielle sont des faits physiques ou neurophysiologiques. Cette seconde option paraît parfaitement plausible car même si j'ai souligné que, selon la conception componentielle, il convient de distinguer le processus comportemental du mouvement physique qui est un constituant du comportement, le comportement n'en est pas moins un processus physique complexe composé d'un état physique interne et d'un mouvement physique. Mais dans ce cas, ne sommes-nous pas conduits à admettre qu'il existe des faits physiques ou neurophysiologiques qui n'ont pas de causes physiques ou neurophysiologiques?

Si on écarte la première option (c'est-à-dire si on rejette l'idée que les faits constitutifs du comportement sont des faits non physiques ou non neurophy-

45. Voir Kim, 1990, 1991.

46. Je suis ici la terminologie de Kim (1990, 1991) qui est elle-même fidèle au sous-titre de l'ouvrage de Dretske, 1988, « Les raisons dans un monde de causes » : une explication par des raisons est une explication qui invoque la propriété sémantique d'un état interne d'un organisme.

siologiques) et si on retient la seconde option, on se retrouve confronté au dilemme auquel était confronté, selon Kim⁴⁷, le partisan du matérialisme non réductionniste, à la seconde étape de l'argument de l'exclusion explicative, lorsqu'il est mis en présence d'un cas de causalité « descendante ». Dans ce cas, l'exemplification d'une propriété sémantique *M* est reliée par une chaîne causale à l'exemplification d'une propriété physique *P**. Comme le partisan du matérialisme non réductionniste doit admettre l'existence d'une base physique ou neurophysiologique *P* de *M* (dont *M* dépend systématiquement), il ne peut octroyer une efficacité causale à la propriété sémantique *M* que s'il consent à enfreindre le principe de la clôture causale du domaine physique et à supposer qu'un phénomène physique ou neurophysiologique peut avoir une cause non physique ou non neurophysiologique.

Le partisan du dualisme des *explananda* et de la conception compositionnelle du comportement doit donc choisir entre deux options : enfreindre le principe de la clôture causale du domaine physique ou renoncer à la composante antiréductionniste du réalisme intentionnel. Comme le suggère Kim⁴⁸, et comme le concède Dretske⁴⁹, la première option semble franchement inadmissible et, pour un partisan du monisme matérialiste, la meilleure stratégie est sans doute la seconde. Mais peut-il choisir la seconde option sans renoncer du coup à la doctrine externaliste? Choisir la seconde option, c'est affirmer que les propriétés sémantiques d'un état interne du cerveau d'un individu sont des propriétés physiques au sens large. Or, qui admet la vérité de l'externalisme reconnaît que les propriétés sémantiques d'une représentation mentale ne dépendent pas systématiquement des propriétés physiques, chimiques et neurophysiologiques intrinsèques du cerveau d'un individu. *A fortiori*, les propriétés sémantiques d'une représentation mentale ne peuvent-elles pas être identiques aux propriétés physiques, chimiques et neurophysiologiques intrinsèques du cerveau d'un individu⁵⁰. L'externalisme n'interdit-il pas la seconde option? Un matérialiste peut-il souscrire à l'externalisme?

Il me semble qu'il peut souscrire à une version « large » de la thèse de l'identité. Les propriétés sémantiques ne sont en effet ni des propriétés intrinsèques du cerveau d'un individu ni des propriétés non physiques. Ce sont des relations entre des états ou des propriétés physiques du cerveau de l'individu et des propriétés physiques exemplifiées dans son environnement. Ce sont donc des relations physiques. S'il admet l'externalisme, un partisan du monisme matérialiste doit concéder que les propriétés sémantiques d'une attitude propositionnelle d'un individu ne sont pas identifiables ou réductibles

47. Voir Kim, 1993a.

48. Voir Kim, 1990, 1991.

49. Voir Dretske, 1991, p. 211, et Dretske, 1995, p. 147-148.

50. Il fait peu de doutes que lorsqu'ils envisageaient (pour la rejeter) l'identification entre les propriétés sémantiques d'une représentation mentale et des propriétés physiques, les adversaires fonctionnalistes du matérialisme réductionniste (dont Putnam et Fodor) pensaient à des propriétés physiques intrinsèques du cerveau d'un individu.

aux propriétés physiques, chimiques ou neurophysiologiques intrinsèques d'un état cérébral de l'individu. Les premières ne dépendent pas systématiquement des secondes. Mais elles dépendent des relations entre les premières et des propriétés physiques exemplifiées dans l'environnement de l'individu. Aussi extrinsèques soient-elles, ces relations n'en sont pas moins physiques pour autant⁵¹.

S'il admet qu'une propriété sémantique n'est autre qu'une relation entre des propriétés neurophysiologiques du cerveau d'un individu et des propriétés physiques exemplifiées dans son environnement, un matérialiste externaliste peut-il encore revendiquer une efficacité causale particulière pour les propriétés sémantiques dans l'explication de la structure du comportement? Je crois que pour répondre au défi de l'exclusion explicative de Kim, le partisan de la dualité des *explananda* et de la conception componentielle du comportement doit recourir à ce que Kim nomme l'« argument de la généralisation⁵² ». Il doit soutenir que les propriétés sémantiques sont dans le même cas que les propriétés non sémantiques des « sciences spéciales » — a propriété chimique d'être une molécule de H₂O, la propriété géologique d'être une montagne ou la propriété biologique d'être un éléphant⁵³. Quoiqu'elles ne soient pas exemplifiées par les particules physiques élémentaires qui constituent le domaine de la physique fondamentale, ces propriétés n'en sont pas moins exemplifiées par des agrégats complexes de particules physiques élémentaires. Ces propriétés sont donc elles-mêmes des propriétés physiques complexes. Comme l'attestent les explications des sciences spéciales, l'efficacité causale de ces propriétés physiques complexes dans des processus chimiques, géologiques et biologiques est indispensable⁵⁴.

Bibliographie

- Churchland, Paul M., 1981. « Eliminative Materialism and the Propositional Attitudes », dans Churchland, Paul, M., *A Neurocomputational Perspective, the Nature of Mind and the Structure of Science*, Cambridge, (Mass.), MIT Press, 1989.
- Davidson, Donald, 1967. « Causal Relations », dans Davidson, Donald, *Essays on Actions and Events*, Oxford, Clarendon Press, 1980.
- Dennett, Daniel, 1971. « Intentional Systems », dans Dennett, Daniel, *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge (Mass.), MIT Press, 1978.

51. Kim, 1987, p. 88, soutient la même chose.

52. Voir Kim, 1998, p. 77. C'est notamment la voie choisie par Fodor, 1989.

53. Comme l'a souligné Millikan, 1993, une entité ne peut être un éléphant à moins d'avoir un pedigree historique phylogénétique. Comme les propriétés sémantiques, la propriété d'être un éléphant est donc une propriété historique.

54. Je dois laisser pour une autre occasion une justification appropriée de ce que Kim, 1998, nomme « l'argument de la généralisation ».

- Dennett, Daniel, 1987. *The Intentional Stance*, Cambridge (Mass.), MIT Press.
- Dretske, Fred, 1988. *Explaining Behavior*, Cambridge (Mass.), MIT Press.
- Dretske, Fred, 1991. « Replies », dans McLaughlin, Brian, dir., *Dretske and his Critics*, Oxford, Blackwell.
- Dretske, Fred, 1995. « Reply : Causal Relevance and Explanatory Exclusion », dans MacDonald, Cynthia & MacDonald, Graham, dir., *Philosophy of Psychology*, Oxford, Blackwell.
- Dretske, Fred, 1998. « Minds, Machines, and Money : What Really Explains Behavior », dans Bransen, Jan & Cuypers, Stefaan, E., dir., *Human Action, Deliberation and Causation*, Dordrecht, Kluwer Academic Publishers.
- Field, Hartry, 1972. « Tarski's Theory of Truth », dans Platts, Mark, dir., *Reference, Truth and Reality*, Londres, Routledge & Kegan Paul.
- Fodor, Jerry, 1987. *Psychosemantics*, Cambridge (Mass.), MIT Press.
- Fodor, Jerry, 1989. « Making Mind Matter More », dans Fodor, Jerry, A., *A Theory of Content and Other Essays*, Cambridge (Mass.), MIT Press, 1990.
- Jacob, Pierre, 1997. *What Minds Can Do*, Cambridge, Cambridge University Press, et *Pourquoi les choses ont-elles un sens?*, Paris, Odile Jacob.
- Jacob, Pierre, 1998. « What Can the Semantic Properties of Innate Representations Explain? », dans Bransen, Jan & Cuypers, Stefaan, E., dir., *Human Action, Deliberation and Causation*, Dordrecht, Kluwer Academic Publishers, 1998.
- Jeannerod, Marc, 1997. *The Cognitive Neuroscience of Action*, Oxford, Blackwell.
- Kim, Jaegwon, 1973. « Causation, Nomic Subsumption, and the Concept of Event », dans Kim, Jaegwon (1993b).
- Kim, Jaegwon, 1984. « Concepts of Supervenience », dans Kim, Jaegwon (1993b).
- Kim, Jaegwon, 1987. « “ Strong ” and “ Global ” Supervenience revisited », dans Kim, Jaegwon (1993b).
- Kim, Jaegwon, 1990. « Explanatory Exclusion and the Problem of Mental Causation », dans MacDonald, Cynthia & MacDonald, Graham, dir., *Philosophy of Psychology*, Oxford, Blackwell, 1995.
- Kim, Jaegwon, 1991. « Dretske on How Reasons Explain Behavior », dans Kim, Jaegwon (1993b).
- Kim, Jaegwon, 1993a. « The Nonreductivist's Troubles With Mental Causation », dans Kim (1993b).
- Kim, Jaegwon, 1993b. *Supervenience and Mind, Selected Philosophical Essays*, Cambridge, Cambridge University Press.
- Kim, Jaegwon, 1998. *Mind in a Physical World : An Essay on the Mind-Body Problem and Mental Causation*, Cambridge (Mass.), MIT Press.
- McDowell, John, 1994. *Mind and World* : Cambridge (Mass.), Harvard University Press.

- Millikan, Ruth Garrett, 1993. *White Queen Psychology and Other Essays for Alice*, Cambridge (Mass.), MIT Press.
- Neander, Karen, 1988. « What Does Natural Selection Explain? Correction to Sober », *Philosophy of Science*, 55, p. 422-426.
- Neander, Karen, 1995. « Pruning the Tree of Life », *The British Journal for the Philosophy of Science*, 46, (1), p. 59-80.
- Salmon, Wesley, 1984. *Scientific Explanation and the Causal Structure of the World*, Princeton, Princeton University Press.
- Searle, John, 1992. *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press.
- Slors, Marc, 1998. « Two Claims That Can Save a Nonreductive Account of Mental Causation », in Bransen, Jan & Cuypers, Stefaan, E., dir., *Human Action, Deliberation and Causation*, Dordrecht, Kluwer Academic Publishers.
- Sober, Elliott, 1984. *The Nature of Selection*, Cambridge (Mass.), MIT Press.
- Stich, Steven, 1983. *From Folk Psychology to Cognitive Science*, Cambridge (Mass.), MIT Press.

Héritabilité causale et propriétés émergentes

Brian Garrett

Le matérialisme contemporain

Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004924ar>

DOI : <https://doi.org/10.7202/004924ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Garrett, B. (2000). Héritabilité causale et propriétés émergentes. *Philosophiques*, 27 (1), 139–159. <https://doi.org/10.7202/004924ar>

Résumé de l'article

Sur la base de ce qu'il a appelé « le principe d'héritabilité causale », Jaegwon Kim a soutenu que les propriétés réalisables de façons multiples ne constituent pas des sortes causales scientifiques. Mon principal objectif est de répondre aux arguments de Kim contre le physicalisme non réductionniste. Je défends l'idée qu'il existe plus de pouvoirs causaux que les seuls pouvoirs causaux physiques. Cela n'a rien de surprenant puisqu'il existe plus de particuliers que le nombre total de particules physiques fondamentales. Et la réflexion sur la nature des individus, et plus spécifiquement sur leur capacité à préserver leur identité à travers le changement ou le remplacement de leurs parties, indique que les individus ont des pouvoirs causaux distincts de ceux des particuliers physiques qui les constituent. Je soutiens que si cela est plausible, alors l'abandon du principe d'héritabilité causale l'est tout autant.

Héritabilité causale et propriétés émergentes

BRIAN GARRETT

bgarrett@yorku.ca

Department of Philosophy

York University, Canada

RÉSUMÉ. — Sur la base de ce qu'il a appelé « le principe d'héritabilité causale », Jaegwon Kim a soutenu que les propriétés réalisables de façons multiples ne constituent pas des sortes causales scientifiques. Mon principal objectif est de répondre aux arguments de Kim contre le physicalisme non réductionniste. Je défends l'idée qu'il existe plus de pouvoirs causaux que les seuls pouvoirs causaux physiques. Cela n'a rien de surprenant puisqu'il existe plus de particuliers que le nombre total de particules physiques fondamentales. Et la réflexion sur la nature des individus, et plus spécifiquement sur leur capacité à préserver leur identité à travers le changement ou le remplacement de leurs parties, indique que les individus ont des pouvoirs causaux distincts de ceux des particuliers physiques qui les constituent. Je soutiens que si cela est plausible, alors l'abandon du principe d'héritabilité causale l'est tout autant.

ABSTRACT. — Jaegwon Kim has recently argued, using what he calls “The Causal Inheritance Principle”, that multiply realizable properties are not causal, scientific kinds. My primary concern is to reply to Kim’s arguments against nonreductive physicalism. I shall defend the idea that there are, indeed, more causal powers than the physical. But this should come as no surprise, since there are more individuals than the total number of fundamental physical particles. Reflections on the nature of individuals, specifically, on their ability to survive through change or replacement of their parts, indicates that individuals have causal powers nonidentical with the causal powers of the physical individuals that constitute them. I claim that, if this is plausible, the rejection of The Causal Inheritance Principle is also plausible.

1. Introduction : épiphénoménisme ou pouvoirs causaux émergents?

Le physicalisme non réductionniste pose que les particuliers, qu'il s'agisse d'objets ou d'événements, sont des entités physiques, tout en niant que chaque aspect ou propriété du monde soit identique à un aspect ou à une propriété physique du monde. Cette thèse est appuyée, notamment, par le fonctionnalisme qui veut que les propriétés mentales soient des propriétés fonctionnelles distinctes des propriétés physiques qui les réalisent de façon contingente. Récemment, le physicalisme non réductionniste a été sérieusement critiqué, et l'une des principales difficultés qui lui a été adressée est celle de savoir si les propriétés survenantes irréductibles¹ ont un rôle causal dans la

1. J'entends la notion de survenance (*supervenience*) en un sens qui implique une relation de dépendance et une relation de réalisabilité multiple : (D) Nécessairement, pour tout x et toute

nature ou si elles ne sont que de simples épiphénomènes, sortes d'appendices aux processus physiques que les matérialistes tiennent pour causalement responsables du changement. Pour plusieurs, la conséquence épiphénoméniste semble inacceptable dans la mesure où elle minerait certaines conceptions plausibles de l'action intentionnelle, de la connaissance et de la responsabilité. Mais si ces propriétés irréductibles (émergentes?) ont réellement une efficacité causale, alors il semble que cela reviendrait à accepter l'existence de pouvoirs causaux qui échappent à la théorie physique, ce qui violerait le principe de clôture du domaine physique, à savoir que chaque changement physique a une explication physique complète, ou une cause entièrement physique. Il semble donc que le physicalisme non réductionniste soit confronté à un dilemme : ou bien les propriétés irréductibles sont causalement inertes et ne contribuent en rien à l'explication des changements physiques, ou bien elles sont causalement efficaces et la théorie physique est incomplète.

Dans cet article, je discute deux arguments de Jaegwon Kim, le plus tenace des critiques du physicalisme non réductionniste. Selon Kim, les deux termes du dilemme devraient être rejetés. Il propose de revenir plutôt à une forme de physicalisme réductionniste, à savoir la thèse de l'identité type-type. Dans ce qui suit, je soutiens que le point de vue de Kim nous engage à une ontologie très peu plausible, et qu'il a certaines implications gênantes relativement à la question de l'identité des personnes. Mais mon principal objectif est de répondre à certains arguments importants que Kim a dirigés contre la thèse du physicalisme non réductionniste. Cependant, plutôt que de défendre cette thèse telle qu'elle est généralement entendue, c'est-à-dire comme impliquant la thèse de l'identité token-token, mes arguments en suggèrent une version qui rejette cette autre thèse, à laquelle le physicalisme non réductionniste est généralement associé. J'endosse le second terme du dilemme et défends l'idée qu'il existe plus de pouvoirs causaux que ce que nous fournit le monde physique². Or cela n'a rien de surprenant, dans la mesure où il est tout à fait plausible qu'il existe plus de particuliers que le nombre total de particules physiques fondamentales plus les sommes méréologiques que constituent les agrégats physiques. Si un particulier, à un temps donné, n'est pas identique à la somme méréologique de ses parties physiques constituantes³, alors la différence entre le particulier et celle-ci doit tenir à certaines propriétés. Autrement dit, si deux particuliers qui ont toutes leurs

propriété *M* de *x*, *x* a une propriété physique *P* telle que nécessairement tout *x* qui a la propriété *P* a la propriété *M* ; (*RM*) Nécessairement, pour toute propriété mentale *M* et pour toute propriété physique *P*, si *P* nécessite *M*, il est possible qu'un objet possède *M* mais ne possède pas *P*.

2. Kim distingue deux thèses : l'émergentisme et la thèse de la réalisabilité physique (*physical realizationism*). Et sur la base du principe de l'héritabilité causale, il soutient que cette dernière serait une thèse moniste. Mon principal argument est que la réalisabilité physique n'est pas moniste, car le principe de l'héritabilité causale est erroné.

3. Cela suppose le rejet du principe d'extensionnalité méréologique qu'on peut exprimer de la façon suivante : Pour tout *x*, pour tout *y* et pour toute partie *p*, si certaines parties composent

parties physiques en commun sont différents, ce doit être en vertu d'une différence dans certaines de leurs propriétés. Et, à mon avis, ces propriétés sont des propriétés causales.

La réflexion sur la nature des individus, et plus spécifiquement sur leur capacité à préserver leur identité à travers le changement ou le remplacement de certaines de leurs parties, indique que les pouvoirs causaux de certains individus (à un temps t), ne sont pas identiques aux pouvoirs causaux de leurs parties physiques constituantes considérées dans leur totalité (au temps t). Considérez par exemple la célèbre statue de Myron, le Discobole, et la masse de bronze qui le compose. Si le Discobole peut survivre au remplacement de certaines de ses parties, alors que la masse de bronze ne le peut pas, ou si la masse de bronze peut changer de forme de façon pratiquement indéfinie tout en préservant son identité, alors que le Discobole ne le peut pas, alors ce serait une raison de nier que le Discobole est identique à la masse de bronze qui le compose (au temps t). Cette dernière compose le Discobole mais la composition n'est pas l'identité — du moins selon ce point de vue qui nie le principe d'extensionnalité méréologique. Cet argument métaphysique est connu et, bien qu'il porte à controverse, c'est un argument que j'accepte. Je veux soutenir que ces individus sont numériquement distincts parce qu'ils possèdent des propriétés différentes, et que certaines de ces propriétés sont justement des propriétés causalement pertinentes ou efficaces. Si le Discobole et la masse de bronze qui le compose ont (au temps t) exactement les mêmes propriétés physiques, alors leur différence doit s'expliquer par une différence de propriétés qui ne sont pas physiques mais qui ne sont pas pour autant épiphénoménales. Voilà donc une problématique où se rencontrent la métaphysique de l'esprit et celle du changement, de la composition et de l'identité.

Mais, qu'en est-il du principe de clôture de la théorie physique qui veut que tout changement physique ait une explication ou une cause physique? Selon mon point de vue, l'existence d'objets qui sont *composés* de sommes d'objets physiques sans leur être identiques ne viole pas, à strictement parler, le principe de clôture du domaine physique. Cela permet simplement de disposer de ressources ontologiques *additionnelles* à celles dont dispose la théorie physique, ressources qui peuvent servir dans des explications causales additionnelles. Si nous acceptons dans notre ontologie des propriétés causales distinctes des propriétés physiques, nous ne devons pas pour autant nier que chaque événement physique, ou chaque exemplification physique de propriétés, ait une cause physique ou une explication physique. Mais cela suggère qu'il existe, en plus des explications physiques, d'autres explications

l'objet x au temps t et ces mêmes parties composent l'objet y au temps t , alors $x = y$. Pour des arguments récents contre ce principe, voir Thomson, J., « The Statue and the Clay », *Nous*, 1998, p. 149-172 ; Baker, L. R., « Why Constitution Is Not Identity », *Journal of Philosophy*, 94, 1997, p. 599-621 ; Doepke, F., « Spatially Coinciding Objects », *Ratio*, 24, 1982.

causales de certains événements physiques : des explications qui réfèrent à des propriétés et à des événements causaux non physiques. À mon avis, l'existence de propriétés causales non physiques est compatible avec le principe de clôture de la théorie physique dans le sens suivant. D'abord, je ne nie pas que chaque événement physique ou chaque exemplification de propriétés ait une cause physique, mais l'existence de propriétés causales qui ne sont pas des propriétés physiques est incompatible avec une doctrine différente, bien que similaire, que j'appelle « la clôture causale de la théorie physique » et selon laquelle il n'existerait pas de propriétés ou d'événements causaux autres que les propriétés ou les événements physiques. C'est cette doctrine que je considère être fausse⁴. Les événements mentaux ne sont pas numériquement identiques aux événements physiques qui les composent, mais ils ont des relations causales avec d'autres événements physiques, tout comme les événements physiques qui composent un événement mental ont des relations causales avec d'autres événements⁵.

Évidemment, nous nous trouvons sur un terrain métaphysique glissant, car nous jonglons simultanément avec les notions de propriétés, d'objets, et d'événements. Ma discussion présuppose que les événements sont des particuliers et que, en tant que tel, il n'est pas incohérent de se poser des questions quant à leur composition par d'autres événements. Le point de vue que je tente de défendre dans ce qui suit est donc le suivant : les événements physiques composent les événements mentaux mais, de façon analogue au Disco-boule et à la masse de bronze qui le compose, la relation de composition n'est pas la relation d'identité. L'événement mental et les événements physiques qui le composent ont en commun exactement les mêmes propriétés physiques, mais l'événement mental n'en est pas moins distinct de la somme des événements physiques qui le composent. De plus, les propriétés qui distinguent l'événement mental de la somme des propriétés physiques qui le composent doivent être entendues comme étant causalement pertinentes.

Pour défendre ce point de vue, je dois montrer (i) que les arguments à l'appui d'une telle différence d'identité numérique d'objets constitués des mêmes parties physiques sont plausibles, (ii) que les propriétés en vertu desquelles un objet est différent de ses parties physiques constituantes ne sont pas des propriétés physiques, et (iii) que ces propriétés sont causalement pertinentes. Mais avant de développer ce point de vue, il est utile de montrer clairement en quoi il est différent du physicalisme de Kim. Aussi, dans la section suivante,

4. Le physicalisme non réductionniste que j'endorsse rejette l'identité token-token, mais il admet que les événements mentaux ont des propriétés physiques (ils ont des constituants physiques). Bien qu'ils ne soient pas numériquement identiques à des événements physiques, ils n'en ont pas moins des propriétés physiques. C'est en ce sens que mon point de vue est compatible avec la clôture du domaine physique.

5. Ce point de vue entraîne donc une forme de surdétermination. Mais celle-ci me semble acceptable si les causes numériquement distinctes sont dans la relation métaphysique de la partie au tout. Voir Garrett, B., « Non-Epiphenomenal Event Dualism », manuscrit.

je discute l'attaque de Kim contre le physicalisme non réductionniste et contre le fonctionnalisme. Je propose ensuite, dans une autre section, une critique du point de vue de Kim ainsi qu'un diagnostic des raisons qui l'ont mené à ses conclusions réductionnistes. Finalement, dans la dernière section, je tente de montrer en quoi les questions concernant la composition et l'identité des objets à travers le temps permettent d'éclairer l'argument de Kim. Si nous souscrivons à la conception métaphysique que je viens de décrire brièvement, nous pouvons rejeter le principe même sur lequel repose l'argument de Kim.

2. L'argument de Kim contre le physicalisme non réductionniste

Dans un texte récent⁶, Kim critique l'interprétation métaphysique du fonctionnalisme qui est largement admise par les tenants du physicalisme non réductionniste. Les propriétés fonctionnelles sont généralement conçues comme étant des propriétés de deuxième ordre. Dire qu'un système *S* a une certaine propriété fonctionnelle *F* revient à dire que *S* a certaines propriétés de premier ordre *P1* ou *P2* ou *P3*, etc., telles que le fait d'avoir l'une ou l'autre de ces dernières est causalement suffisant pour produire certains effets. Comme Kim le note :

F est une propriété de deuxième ordre définie sur un ensemble *B* de propriétés de base si et seulement si *F* est la propriété d'avoir une certaine propriété *P* appartenant à l'ensemble *B*, telle que *D(P)*, où *D* spécifie une certaine condition sur les membres de *B*⁷.

Si *F* est une propriété fonctionnelle, alors la condition *D* spécifie un certain rôle fonctionnel. Ainsi, par exemple, la solubilité est une propriété fonctionnelle de deuxième ordre. Pour qu'une substance, disons le sel, soit soluble, il doit exister une certaine propriété de premier ordre, disons la propriété d'être composé chimiquement de NaCl, telle que le fait d'avoir cette constitution chimique suffit à ce que le sel se dissolve s'il est placé dans un certain liquide non saturé approprié.

Il a été largement admis que le rôle fonctionnel est une propriété distincte des propriétés qui le réalisent. Par exemple, la solubilité est une propriété distincte de la propriété qui caractérise la constitution chimique du sel ou de celle qui caractérise le sucre : chacune de ces dernières réalise la solubilité. Cette idée a permis à certains fonctionnalistes comme Hilary Putnam, du moins à l'époque où il était encore fonctionnaliste, et Jerry Fodor de soutenir un certain dualisme des propriétés, puisqu'ils croyaient que les propriétés fonctionnelles ne sont pas des propriétés physiques, compte tenu de la thèse de la réalisabilité multiple qui veut qu'une même propriété fonctionnelle puisse avoir des réali-

6. Kim, J., « The Mind-Body Problem : Taking Stock After Forty Years », dans Tomberlin, J., dir., *Philosophical Perspectives*, vol. 11, Boston, Blackwell, 1997.

7. *Ibid.*, p. 195.

sations physiquement hétérogènes, comme le sel et le sucre dans l'exemple de la solubilité⁸. De plus, comme les propriétés fonctionnelles sont essentiellement caractérisées en faisant référence à des propriétés de premier ordre qui ont un certain rôle causal, le fonctionnalisme est apparu comme une façon de résoudre facilement les problèmes traditionnels concernant l'efficacité causale des propriétés mentales, désormais considérées comme des propriétés fonctionnelles. En effet, si les propriétés mentales sont des propriétés fonctionnelles de deuxième ordre, alors elles seraient causalement pertinentes, par définition, puisque les propriétés fonctionnelles requièrent l'existence de propriétés de premier ordre qui satisfont un certain rôle causal⁹. Finalement, un des attraits du fonctionnalisme en philosophie de l'esprit est de s'inscrire tout naturellement dans une certaine ontologie fonctionnaliste. Autrement dit, cette approche fournirait le schéma d'une ontologie générale : les propriétés fonctionnelles, comme William Lycan l'a suggéré, se retrouveraient à divers niveaux dans une organisation hiérarchique de la nature :

Si nous considérons la nature comme étant hiérarchiquement organisée de cette façon, alors la distinction « fonction/structure » devient relative : quelque chose est un rôle par opposition à l'occupant d'un rôle, un état fonctionnel par opposition à une réalisation de cet état, ou vice versa, modulo un certain niveau spécifique de la nature¹⁰.

Ainsi, la nature comprendrait une hiérarchie de propriétés ontologiquement distinctes et causalement efficaces, et les propriétés se situant à un certain niveau seraient survenantes à celles se situant à un niveau inférieur. Ultimement, toutes seraient survenantes à des propriétés fondamentales, supposément les propriétés physiques de l'univers. La question de savoir si une propriété est ou non une propriété fonctionnelle de deuxième ordre serait une question pragmatique qui dépend de nos intérêts dans divers contextes¹¹. Relativement à la psychologie, les propriétés neurologiques

8. Voir Putnam, H., « Philosophy and Our Mental Life », dans *Mind, Language and Reality : Philosophical Papers*, vol. 2, Cambridge (Mass.), Cambridge University Press, 1975 ; et Fodor, J., « Special Sciences : Still Autonomous After All These Years », dans Tomberlin, J., dir., *Philosophical Perspectives*.

9. Cependant, Kim lui-même ainsi que Ned Block ont soutenu que la propriété d'avoir une propriété causalement efficace n'est pas elle-même une propriété causalement efficace ; voir Block, N., « Can the Mind Change the World », dans MacDonald, C. et MacDonald, G., dir., *Philosophy of Psychology. Debates on Psychological Explanation*, Oxford, Blackwell, 1995.

10. Lycan, W., « The Continuity of Levels of Nature », dans Lycan, W., dir., *Mind and Cognition*, Cambridge (Mass.), Blackwell, 1990, p. 78.

11. Cette façon de parler pourrait laisser croire que les propriétés fonctionnelles et les propriétés de deuxième ordre sont une seule et même chose, mais ce n'est pas le cas, comme l'a noté Fodor dans « Special Sciences : Still Autonomous After All These Years ». Le jade est une propriété de deuxième ordre car ses exemplifications requièrent l'exemplification de propriétés réalisantes, mais elle n'est pas fonctionnelle car elle requiert des propriétés réalisantes très spécifiques et non pas simplement toute propriété réalisante qui capturerait les causes et les effets macrophysiques du jade.

seraient des propriétés réalisantes de premier ordre, mais relativement à la sociologie, les propriétés psychologiques seraient elles-mêmes des propriétés réalisantes de premier ordre.

Jaegwon Kim a récemment critiqué l'approche fonctionnaliste en s'appuyant sur divers arguments. Le premier, et peut-être le plus important, est dirigé contre les arguments de réalisabilité multiple (RM) qui appuient l'indépendance ontologique des propriétés fonctionnelles. Kim tente de bloquer ces arguments dans le but de montrer que si les propriétés survenantes peuvent être réalisées de multiples façons, alors elles ne peuvent être des propriétés scientifiques ou naturelles légitimes, ce qui serait une mauvaise nouvelle pour les tenants de l'irréductibilité ontologique des sciences « spéciales ». À tout le moins, ce serait une mauvaise nouvelle pour ceux qui, comme Jerry Fodor, soutiennent que les propriétés irréductibles pourraient être étudiées systématiquement dans le contexte de lois scientifiques exprimant leurs pouvoirs causaux¹².

Le deuxième argument de Kim s'attaque à la définition des propriétés fonctionnelles. Si celles-ci doivent être conçues comme des propriétés de deuxième ordre, alors nous n'aurions pas de motifs pour faire des affirmations ontologiques substantielles à leur propos. Si une propriété fonctionnelle est une propriété de deuxième ordre, c'est-à-dire la propriété d'avoir certaines propriétés de premier ordre satisfaisant certains rôles causaux, alors, nous dit Kim, une telle propriété ne serait pas réellement distincte des propriétés de premier ordre qui la réalisent. Nous discutons cet argument de façon plus détaillée dans ce qui suit.

La thèse de la réalisabilité multiple a été utilisée par Putnam et par Fodor pour soutenir que les propriétés physiques et mentales sont ontologiquement distinctes. L'idée est tout simplement qu'une propriété mentale serait réalisée par une longue disjonction, voire une disjonction infinie, de sortes physiques. Or une longue disjonction de sortes physiques n'est pas elle-même une sorte physique, car le prédicat disjonctif qui dénote cette « propriété » n'aurait pas le caractère de projectivité qui est un des critères généralement admis pour qu'un prédicat dénote une sorte physique. Ainsi, la propriété mentale ne pourrait être identifiée à une propriété physique, puisque cette dernière n'existe tout simplement pas. Kim renverse cet argument et il l'utilise pour montrer justement que si les propriétés mentales peuvent être réalisées de multiples façons, alors elles ne constituent tout simplement pas des sortes naturelles, dans la mesure où les « propriétés » disjonctives qui leurs sont coextensives ne sont pas des sortes naturelles. Le point central de l'argument de Kim est que les propriétés survenantes doivent hériter leurs

12. Fodor, J., « Special Sciences, or the Disunity of Science as a Working Hypothesis », dans Block, N., dir., *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, p. 120-133.

pouvoirs causaux, et par conséquent leur statut taxinomique, des propriétés physiques qui les réalisent. Considérons les trois principes suivants :

Principe de réalisabilité multiple (RM) : Nécessairement, pour toute propriété mentale M et pour toute propriété physique P qui nécessite M , il est possible qu'un objet possède M mais ne possède pas P .

Principe d'héritabilité causale (HC) : Si la propriété M est réalisée dans un système, au temps t , par une propriété physique P , les pouvoirs causaux de cette occurrence particulière de M (disons M_i) sont identiques aux pouvoirs causaux de cette occurrence particulière de P (disons P_i).

Principe d'individuation causale des sortes naturelles (ICS) : Les sortes naturelles en science sont individuées sur la base de leurs pouvoirs causaux ; c'est-à-dire que les objets et les événements sont d'une même sorte naturelle ou partagent une même propriété seulement s'ils ont des pouvoirs causaux similaires¹³.

Il est utile de commenter brièvement chacun de ces principes. Selon (RM), les propriétés mentales sont réalisées, et peuvent être réalisées, par des propriétés physiques hétérogènes. Les exemples standards incluent la possibilité que le cerveau d'un extraterrestre, un « cerveau » informatique de silicone et un cerveau humain puissent réaliser la même propriété psychologique, par exemple la propriété de croire-que-Leibniz-était-brillant. Le troisième principe illustre le point de vue de Kim à propos des sortes naturelles. En fait, il exprime sans doute deux thèses différentes, à savoir : (ICS.i) les sortes naturelles en science sont des sortes causales, et (ICS.ii) si $K = K^*$, alors $K(a)$ doit être causalement similaire à $K^*(b)$ ¹⁴. (ICS.i) paraît douteux s'il est entendu comme une généralisation universelle, puisque Kim ne veut probablement pas dire que les sortes en mathématiques ou dans les sciences abstraites sont des sortes causales¹⁵. Cela laisse ouverte la possibilité que la psychologie elle-même soit une science abstraite, c'est-à-dire une science qui s'intéresserait essentiellement à la rationalité, et dont l'objectif serait de produire des rationalisations plutôt que des généralisations causales. Peut-être. Mais une telle objection à (ICS.i) ne peut être utilisée dans le cas d'autres sciences « spéciales » qui s'intéressent également à des propriétés survenantes. Il est en effet difficile de croire que les propriétés géologiques ou biologiques sont des propriétés abstraites causalement inefficaces. Donc j'accepte (ICS.i).

(ICS.ii) porte sur l'individuation des sortes scientifiques. Cependant, ce principe est plutôt vague : jusqu'à quel point deux occurrences d'une pro-

13. Fodor, J., « Methodological Solipsism », dans *RePresentations. Philosophical Essay on the Foundations of Cognitive Science*, Cambridge (Mass.), MIT Press, 1981, p. 225-253.

14. Où « a » et « b » réfèrent à des particuliers.

15. Mais pourquoi pas? Si les propriétés sont des sortes causales, dans la mesure où leurs exemplifications en sont des réalisations physiques, alors puisque les propriétés mathématiques sont exemplifiées par des états physiques, celles-ci seraient également des sortes causales. Mais peut-être qu'il ne suffit pas qu'une propriété ait des réalisations physiques pour qu'elle constitue une sorte causale ; quelque chose de plus est peut-être requis.

priété (dans des objets distincts) doivent-elles être causalement similaires pour qu'elles soient des occurrences distinctes d'une même propriété? Les jugements de similarité et d'hétérogénéité sont évidemment relatifs à un arrière-plan contextuel, et cela vaut également pour les jugements de similarité causale. Le problème n'est pas tant que (ICS.ii) soit erroné (je l'endosse également), mais qu'il puisse être interprété de telle sorte qu'il constituerait une pétition de principe contre le physicalisme non réductionniste. Si on exige que les jugements de similarité causale utilisent le vocabulaire physicaliste, alors le principe devient douteux : les tenants de la thèse selon laquelle il existe des propriétés causalement efficaces mais irréductibles à des propriétés physiques n'accepteraient pas cette interprétation. Si aucune contrainte de ce type ne restreint la notion de similarité causale, alors le principe semble difficilement contestable. Cependant, comme je le montre dans la section suivante, l'argument de Kim devient alors douteux.

Finalement, selon le principe d'héritabilité causale, les occurrences de propriétés réalisées de multiples façons hériteraient leurs propriétés causales directement des occurrences de propriétés physiques qui les réalisent. Ce principe est bien évidemment une conséquence directe de la thèse de l'identité token-token. Si une occurrence de *Mi*, la croyance-que-Leibniz-était-brillant, est identique à un certain événement neuronal, disons *Ni*, alors cette occurrence mentale doit avoir exactement les propriétés causales de l'événement neuronal auquel elle est identique! Aussi, dans la mesure où on est enclin à accepter la thèse de l'identité token-token, on devrait accepter (HC), puisque selon ce point de vue moniste les occurrences de propriétés mentales sont identiques à des occurrences de propriétés physiques (même si les propriétés mentales elles-mêmes, ou types mentaux, ne sont pas identiques à des propriétés physiques). Ce point de vue s'accommode bien du monisme anomal de Davidson, pour qui les événements mentaux sont identiques à des événements physiques qui sont eux-mêmes les causes du comportement¹⁶. Bref, le principe d'héritabilité causale semble être endossé par les tenants du physicalisme non réductionniste, et par conséquent il devrait être sacro-saint.

Or ces trois principes semblent avoir des conséquences dévastatrices pour l'ontologie du physicalisme non réductionniste. Considérons la propriété mentale *M*, disons croire-que-*p*, et deux occurrences de cette propriété, *Mi* et *Mii*, réalisées chez deux individus, disons un extraterrestre et un humain. Selon (RM), *M* est réalisée par des événements physiques distincts, qui exemplifient également deux propriétés physiques *P1* et *P2*. Selon l'héritabilité causale, *Mi* et *Mii* auraient respectivement les mêmes propriétés causales que les deux événements physiques (*P1i* et *P2i*) qui les réalisent. Or, il semble y avoir un sérieux problème lorsque nous considérons le troisième principe. En effet, si nous supposons que *Mi* et *Mii* ont, respectivement, les mêmes propriétés causales que

16. Voir Davidson, D., « Mental Events », dans Foster, L. et Swanson, J. W., dir., *Experience and Theory*, Amherst, The University of Massachusetts Press, 1970.

P1i et *P2i*, la question se pose de savoir comment *Mi* et *Mii* devraient être individuées, c'est-à-dire sous quelle(s) sorte(s) naturelle(s) nous devrions les catégoriser. *Mi* et *Mii* auraient des pouvoirs causaux différents puisqu'elles auraient exactement les pouvoirs causaux des occurrences de propriétés physiques distinctes (*P1* et *P2*). Ainsi, compte tenu du principe d'individuation, *Mi* et *Mii* ne seraient pas deux occurrences d'une même sorte scientifique *M*, car les deux réalisations de *M* sont *causalement* diverses et, selon (ICS), elles ne pourraient être catégorisées comme appartenant à la même sorte naturelle. Donc, *M* ne peut constituer une sorte scientifique. Si (RM) vaut généralement pour les propriétés mentales, géologiques, météorologiques, etc., alors celles-ci ne constitueraient tout simplement pas des sortes scientifiques! Les prédicats utilisés dans ces disciplines exprimeraient simplement des concepts plutôt que des propriétés naturelles. Comme le note Kim :

Le raisonnement est simple : les occurrences de *M* qui sont réalisées par une même base physique doivent être groupées sous une même sorte puisque, comme nous le supposons, la base physique constitue une sorte causale ; et les occurrences de *M* qui ont des bases physiques différentes doivent être groupées sous des sortes différentes puisque, comme nous le supposons toujours, ces bases de réalisation physiques distinctes constituent des sortes causales distinctes. Étant donné que les sortes mentales sont réalisées par des sortes physiques causalement distinctes, il en résulte donc que les sortes mentales ne sont pas des sortes causales, et ainsi elles ne constituent pas des sortes scientifiques adéquates¹⁷.

L'exemple utilisé par Kim illustre bien son argument. Comme nous le savons, le jade est coextensif avec la jadéite et la néphrite, qui sont des sortes naturelles distinctes, ce que révèle leur description moléculaire. Une loi à propos du jade, comme par exemple « Le jade est vert », est nomologiquement coextensive à deux lois : « la jadéite est verte » et « la néphrite est verte ». Ainsi, « Le jade est vert » n'est pas une généralisation nomologique car le terme « jade » n'est pas projectible, puisque le jade est tantôt réalisé par la jadéite, tantôt par la néphrite, lesquelles sont chimiquement hétérogènes. Si nous essayons de confirmer l'énoncé général « Le jade est vert » et que tous nos échantillons sont des échantillons de jadéite, alors il ne serait pas légitime d'inférer sur cette base inductive que le prochain échantillon de jade, disons un échantillon de néphrite, sera probablement vert. Un tel saut inductif ne serait pas fondé compte tenu de la différence moléculaire entre la jadéite et la néphrite.

Si (RM) est vrai, alors les propriétés réalisées de multiples façons ne sont pas des sortes naturelles, tout comme le jade n'est pas une sorte naturelle compte tenu de ses réalisations hétérogènes. Cependant, le prédicat « jade » n'en est pas moins utile pour nos explications de la vie de tous les jours. Ainsi, selon Kim, ce prédicat exprimerait un concept plutôt qu'une propriété ayant une efficacité causale. Pour reprendre l'exemple que nous

17. Kim, J., « Multiple Realization and the Metaphysics of Reduction », *Philosophy and Phenomenological Research*, 52, 1992, p. 19.

avons déjà utilisé, croire-que-Leibniz-était-brillant ne serait pas une sorte naturelle, car cette propriété est réalisable de multiples façons par des propriétés physiques différentes, par exemple par des machines de silicone, par des extraterrestres, par des humains et ainsi de suite.

Le deuxième argument de Kim procède de la façon suivante. Si les propriétés fonctionnelles sont des propriétés de deuxième ordre, alors elles ne sont rien de plus que les propriétés de premier ordre qui les réalisent. Une propriété de deuxième ordre est formée par quantification existentielle sur un domaine de propriétés de premier ordre. La solubilité consiste dans le fait d'avoir une propriété (chimique) de premier ordre qui a certains effets appropriés (si une substance soluble est plongée dans un liquide approprié dans certaines conditions *C*, alors elle se dissout). Le jade est une propriété de deuxième ordre dans le sens où *x* est du jade s'il existe certaines propriétés de base qui satisfont le rôle causal du jade, à savoir la propriété d'être de la jadéite ou celle d'être de la néphrite. Mais comme le note Kim, ce serait pure magie si une telle quantification existentielle sur des propriétés de premier ordre suffisait à générer une nouvelle propriété distincte :

Nous pouvons commencer par reconnaître explicitement qu'en généralisant existentiellement sur un certain ensemble de propriétés, nous n'amenons pas à l'existence un nouvel ensemble de propriétés. Cela serait pure magie : par des opérations logiques sur notre notation, comme la quantification, nous ne pouvons modifier notre ontologie — nous ne pouvons ni la diminuer ni l'augmenter. [...] En quantifiant sur des propriétés, nous ne pouvons créer de nouvelles propriétés, pas plus que nous ne pouvons créer de nouveaux individus en quantifiant sur des individus. Quelqu'un a assassiné Dupont et cet assassin est ou Dumont, ou Durand, ou Tremblay. Ce « quelqu'un » qui a assassiné Dupont n'est pas une autre personne en plus de Dumont, Durand, et Tremblay, et il serait absurde de postuler une personne disjunctive, soit Dumont-ou-Durand-ou-Tremblay qu'on identifierait à l'assassin. Le même raisonnement vaut pour les propriétés de second ordre et leurs réalisations¹⁸.

Tout comme une disjonction de propriétés *P1* ou *P2* ou ... *Pn* n'exprime pas une propriété, une quantification existentielle comme « Il y a une propriété *P* telle que *x* est *P* » n'exprimerait pas non plus une propriété. Le fonctionnalisme, qui veut que les propriétés fonctionnelles soient des propriétés de deuxième ordre formées par quantification existentielle sur des propriétés de premier ordre, ne nous permettrait pas de réifier les propriétés fonctionnelles comme étant de nouvelles propriétés irréductibles. Par conséquent, si les propriétés invoquées dans les sciences spéciales sont bien des propriétés de deuxième ordre, elles ne seraient finalement pas différentes des propriétés de premier ordre qui les réalisent. Si, par exemple, la propriété d'être un glacier n'est rien d'autre que la propriété d'avoir certaines propriétés de premier ordre ayant un certain rôle causal que l'on peut attendre des glaciers — comme celle

18. Kim, « The Mind-Body Problem : Taking Stock After Forty Years », p. 201.

d'être de la glace et celle d'être de la neige —, alors la propriété d'être un glacier ne serait pas une propriété au-delà de ses propriétés réalisantes.

3. Objections à l'argument de Kim

Kim croit qu'il serait moins trompeur de parler de concepts plutôt que de propriétés de deuxième ordre. Il serait préférable, selon lui, d'entendre les propriétés fonctionnelles comme étant des concepts ou des descriptions qui dénotent des propriétés de premier ordre. Si les « propriétés » fonctionnelles sont entendues comme n'étant rien d'autre que des concepts, cela nous permettrait d'en préserver l'utilité conceptuelle tout en faisant une économie ontologique : nous pouvons utiliser de tels concepts fonctionnels tout en restant silencieux à propos des propriétés de premier ordre. Nous pouvons parler de la solubilité d'une substance sans avoir à être plus précis sur les propriétés chimiques qui sont suffisantes pour la solubilité. De tels concepts véhiculent de l'information pertinente dans les contextes où nous les utilisons, information qui pourrait être occultée si nous utilisions plutôt les prédicats exprimant les propriétés de premier ordre. L'approche de Kim est attrayante, principalement parce qu'elle est ontologiquement parcimonieuse : avoir une propriété fonctionnelle n'est rien d'autre que d'avoir une propriété réalisante quelconque. Nous n'avons pas besoin d'empiler les unes par-dessus les autres des propriétés de plusieurs niveaux différents.

Mais la position générale qui résulte de l'approche de Kim est difficile à croire, car si elle est correcte, alors il serait exclu que les propriétés fonctionnelles soient des sortes scientifiques. Plus exactement, il n'existerait pas de propriétés réalisables de multiples façons mais seulement des concepts qui peuvent être appliqués de multiples façons, et la hiérarchie de propriétés fonctionnelles irréductibles s'effondrerait. Les propriétés physiques existent et elles sont fondamentales mais elles ne seraient pas fondamentales au sens où elles seraient constitutives des propriétés des diverses sciences, mais simplement dans le sens où elles détermineraient les conditions d'applicabilité des concepts utilisés dans ces diverses sciences. S'il doit exister dans les sciences spéciales des propriétés ontologiquement robustes, alors celles-ci ne pourraient être réalisées de multiples façons, car elles seraient coextensives à des sortes causales légitimes du domaine physique. Dans la mesure où la plupart des « propriétés » des sciences spéciales sont réalisées de façons multiples, comme Jerry Fodor l'a souligné, l'argument de Kim serait du coup un argument pour l'élimination ontologique de la très vaste majorité des propriétés des sciences spéciales. Dès lors, l'extraterrestre et l'humain qui ont la croyance-que-Leibniz-était-brillant n'auraient plus de propriété en commun, bien que le *concept* « croire que Leibniz est brillant » pourrait leur être appliqué, tout comme le concept « jade » peut être appliqué à deux quantités de jadéite et de néphrite. L'argument de Kim est peu crédible.

Il semble peu plausible qu'il n'existe pas de propriétés des sciences spéciales réalisées de multiples façons. S'il n'existe pas de telles sortes scientifiques, alors que dire des objets qui instancient ces « propriétés »? Si le prédicat « est une montagne » peut être réalisé de façons multiples alors, selon Kim, il n'y aurait pas de propriété causale d'être une montagne. Cela semble avoir pour conséquence qu'à proprement parler, il n'existerait pas un objet d'étude scientifique dont la caractéristique est d'avoir cette propriété. Mais une telle conclusion n'est-elle pas difficile à croire? Quel serait donc cet objet que les géologues ont essayé de comprendre? De plus, cette conclusion est parfaitement générale : il n'y aurait pas d'objets scientifiques individués par des propriétés réalisées de multiples façons (car il n'y a pas de telles propriétés), de sorte que la plupart des objets que nous acceptons dans le discours de tous les jours n'existeraient pas comme objets scientifiques. Mais ici nous devons être prudents : l'argument de Kim n'entraîne pas la conclusion qu'il n'existe pas de montagnes ou qu'il n'existe pas de créatures psychologiques. Par contre, ce qui résulte de son argument, c'est qu'il ne peut y avoir de science causale des propriétés psychologiques ou de science causale de la propriété d'être une montagne.

Si les propriétés psychologiques sont réalisées de multiples façons — disons dans des créatures de silicone et dans des créatures de carbone — alors ces « propriétés » ne seraient pas des sortes causales tombant dans le domaine d'étude de la science psychologique, puisque aucune science psychologique ne pourrait subsumer sous une même catégorie ces deux créatures. Une science psychologique causale pourrait subsumer sous une même catégorie les êtres humains, à condition que les réalisations physiques des propriétés psychologiques des humains ne soient pas hétérogènes. Même si les extraterrestres, les machines à base de silicone et les être humains ne tombent pas tous sous des sortes naturelles de la psychologie humaine, cela n'exclut pas la possibilité de diverses sciences psychologiques, par exemple la psychologie proprement humaine, la psychologie des extraterrestres, et la psychologie des machines. Cependant, il faut bien voir que même si ces diverses sciences portaient toutes le nom de « psychologie », ces sciences ne seraient pas liées entre elles comme différentes familles d'une même espèce, car elles ne s'intéresseraient aucune-ment à des propriétés psychologiques communes à des créatures de diverses sortes. Ce qui est gênant avec un tel point de vue c'est qu'il aurait le désavan-tage de nous faire perdre des généralisations qui paraissent subsumer correc-tement ces diverses créatures, comme Ned Block l'a souligné¹⁹.

Si nous devons accepter le point de vue de Kim, certaines thèses méta-physiques tout à fait plausibles concernant l'identité personnelle et la conti-nuité devraient être abandonnées. Considérons, par exemple, un processus extrêmement lent par lequel les neurones et les connexions neuronales de votre cerveau seraient graduellement remplacés par des mécanismes artificiels capa-

19. Block, N., « Anti-Reductionism Slaps Back », dans Tomberlin, J., dir., *Philosophical Perspectives*, p. 107-132.

bles d'effectuer les mêmes transferts d'énergie et d'information que ceux effectués par vos neurones. Afin d'éviter la dégénérescence prévue de vos cellules nerveuses et de leurs fonctions, il n'est pas inconcevable que vos neurones puissent être remplacés de la sorte par des éléments artificiels ou par un matériau biologique artificiellement « cultivé », éléments qui seraient physiquement différents des neurones d'origine. Mon intuition est qu'un tel processus pourrait préserver intactes nos propriétés psychologiques. Comme le but pratique visé dans un tel scénario est justement de préserver le fonctionnement psychologique de l'individu, c'est uniquement la préservation des effets pertinents pour ce but qui est désirée. Les mécanismes artificiels, qu'ils soient biologiques ou non, pourraient eux-mêmes être constitués de propriétés physiques hétérogènes et ils auraient donc des pouvoirs causaux différents. Ces mécanismes pourraient avoir des effets tout à fait distincts s'ils étaient placés dans d'autres médiums, mais dans le contexte d'un cerveau ils pourraient bien accomplir les mêmes tâches. Le point important est qu'un tel processus de remplacement graduel des neurones pourrait réussir et, si c'était le cas, qu'il produirait un patient qui, du point de vue de la troisième personne, serait en parfaite santé psychologique.

Mais en raison du changement supposé dans la base de réalisation qui transforme une neurophysiologie humaine en l'une ou l'autre de ces « physiologies » artificielles, Kim doit nier que le patient exemplifie les mêmes propriétés psychologiques que celles qu'il exemplifiait avant l'opération. Selon son point de vue, une telle opération devrait échouer en principe si le but est de préserver la psychologie d'origine du patient. Kim pourrait sans doute répondre que même s'il n'y a pas de continuité psychologique avant et après l'opération, il y aurait un isomorphisme fonctionnel. Mais cela équivaldrait à la perte de la personne en cours d'opération et, curieusement, ce serait une perte de propriétés psychologiques que seul un physicien pourrait détecter! À mon avis, même si cette objection n'est peut-être pas dirimante, dans la mesure où les opinions philosophiques à propos de la nature de l'identité personnelle divergent considérablement, elle soulève tout de même une sérieuse difficulté. Kim pourrait sans doute accepter la conséquence selon laquelle des opérations de ce genre ne préserveraient pas l'identité personnelle, puisqu'elles ne préserveraient pas la continuité psychologique. Mais les conséquences épistémologiques d'un tel point de vue sont quelque peu déroutantes. En effet, il semble bien que ni le patient d'une telle opération ni des observateurs extérieurs ne seraient en mesure de savoir qu'il y a eu perte de la personne au cours de l'opération, à moins qu'ils ne connaissent les faits physiques pertinents, ce qui semble épistémiquement contre-intuitif. Par ailleurs, Kim pourrait nier que l'identité personnelle dépende de la continuité et de la connectivité psychologiques²⁰. Peut-être que l'identité personnelle à travers le temps pourrait être préservée en faisant l'économie de

20. Les notions de continuité psychologique et de connectivité psychologique ont été caractérisées par Derek Parfit. Selon Parfit, il y a connectivité psychologique lorsqu'il existe

ces deux caractéristiques. Cependant, je dois admettre que je ne comprends tout simplement pas en quel sens une théorie qui nie la nécessité de la continuité psychologique ou de la connectivité (plutôt que de nier simplement que ce sont là des conditions suffisantes) puisse être une théorie de l'identité des personnes, plutôt qu'une théorie de leur corps.

La position générale de Kim a donc des conséquences qui semblent difficilement acceptables. Qu'a-t-il bien pu se passer? La critique la plus intéressante (et sans doute la plus discutable) consiste à rejeter un ou plusieurs des trois principes sur lesquels repose l'argument de Kim. Dans ce qui suit, je propose de rejeter le principe d'héritabilité causale. Mais revenons d'abord à l'ambiguïté du principe d'individuation causale des sortes naturelles. Selon ce principe « [...] les objets et les événements sont d'une même sorte naturelle, ou partagent une même propriété, seulement s'ils ont des pouvoirs causaux similaires ». Mais si nous considérons deux particuliers quelconques, ceux-ci peuvent toujours être considérés comme ayant des pouvoirs causaux similaires ou hétérogènes, si nous prenons la peine de choisir de façon judicieuse la mesure de similarité. Ainsi, deux exemplifications d'une certaine propriété physique seront causalement hétérogènes relativement à une certaine mesure. Nos jugements de similarité causale dépendent des détails auxquels nous portons attention, c'est-à-dire des effets et des causes qui nous importent dans un contexte particulier.

L'objection est que nous devons toujours décider d'une certaine mesure de similarité avant de pouvoir évaluer la situation. Un tenant de l'efficacité causale des propriétés réalisables de multiples façons ne manquera pas de noter que, par définition, (RM) requiert l'hétérogénéité à un certain niveau de description et l'homogénéité à un autre niveau. Ainsi, les occurrences de propriétés survenantes réalisables de multiples façons ont des effets très similaires lorsque ces occurrences sont catégorisées à l'aide du vocabulaire de la discipline appropriée. On pourrait tenter de résister à cette objection en insistant sur le fait que les propriétés réalisables de multiples façons doivent être des sortes causales, que deux occurrences d'une même sorte doivent être causalement similaires, et qu'il serait correct de dire qu'elles le sont seulement si nous sommes forcés de mesurer leur similarité causale par des standards *physiquement* adéquats. Mais un tel raisonnement semble justement présupposer ce qu'il tente d'établir. Comme je l'ai déjà noté, la notion de « similarité causale » que nous utilisons dans cette discussion doit être neutre (*topic neutral*).

certaines connexions psychologiques directes entre des états mentaux du sujet. De telles connexions directes sont des relations causales. La connectivité psychologique est une question de degré qui dépend du nombre de telles connexions directes. Sur cette base, Parfit caractérise la notion de connectivité *forte* qui, bien qu'elle ne puisse être définie strictement, signifie tout simplement que le degré de connectivité psychologique est « assez grand » pour que le sujet soit une personne. La continuité psychologique est ensuite définie comme une chaîne continue de relations de connectivité *forte*, qui se chevauchent. Voir Parfit, D., *Reasons and Persons*, Oxford, Clarendon Press, p. 206.

On pourrait également être tenté de répliquer en notant que l'occurrence particulière d'une propriété survenante est un événement identique à l'occurrence de la propriété physique qui la réalise et, par conséquent, que le même critère de similarité causale devrait s'appliquer à « chacune ». Comme, selon (RM), les diverses réalisations sont hétérogènes, alors, étant donné l'identité token-token, les propriétés survenantes seraient aussi hétérogènes. Mais cette réplique passe à côté de l'objection. Même si nous concédions cette version du physicalisme non réductionniste — bien que finalement je désire résister à cette interprétation —, nous devrions de toutes façons poser la question de savoir si les deux événements sont ou non causalement similaires. Imaginons, par exemple, deux événements qui sont des cas d'ingestion d'une substance dormitive, dans un cas de l'opium et dans l'autre un sédatif obtenu à la pharmacie. Que ces deux événements soient causalement similaires ou non dépend, encore une fois, des types d'effets et de causes que nous considérons être pertinents. Si toutes les différences étaient pertinentes, alors nous devrions abandonner l'idée qu'une propriété causalement efficace puisse être exemplifiée plus d'une fois. Si nous ne considérons que la similarité de ce qu'on pourrait appeler les micro-effets comme étant pertinente, alors les deux événements pourraient bien être causalement différents, dans la mesure où il s'agit d'ingestion de deux types de sédatifs différents ; mais si la similarité d'effets macroscopiques sur la psychologie humaine est notre critère, ces événements pourraient être classifiés comme étant causalement similaires.

4. Contre le principe d'héritabilité causale

Le principe d'héritabilité causale joue un rôle fondamental dans l'argument de Kim. Le nom de ce principe est particulièrement évocateur. Lorsque nous héritons de quelque chose nous obtenons quelque chose qui nous faisait défaut ; mais lorsque nous héritons, nous n'héritons pas nécessairement de la totalité des biens héréditaires : il peut arriver que nous n'obtenions qu'une partie du magot, le reste pouvant aller à des œuvres de charité ou à d'obscurs cousins oubliés. De plus, être un héritier ne requiert pas que nous soyons au préalable indigents. Or Kim utilise la notion d'héritabilité comme un principe qui n'accepte pas de degré et qui s'applique de façon *tout ou rien*. Les occurrences de propriétés mentales nettoient tout sur leur passage : elles héritent très exactement de tous les pouvoirs causaux des occurrences de propriétés physiques qui les réalisent. J'aimerais suggérer que la métaphore d'héritabilité ne devrait pas être restreinte de la sorte et que le principe devrait être révisé. Comme je l'ai déjà noté, le principe est le suivant :

Principe d'héritabilité causale (HC) : Si la propriété mentale M est réalisée dans un système, au temps t , en vertu d'une base de réalisation physique P , les pouvoirs causaux de cette occurrence particulière de M (disons M_i) sont identiques aux pouvoirs causaux de cette occurrence de P (disons P_i).

Cette thèse est particulièrement plausible pour une certaine version du physicalisme non réductionniste, à savoir la thèse du monisme anomal de Davidson. Si les événements mentaux sont des particuliers concrets et qu'ils sont identiques à des événements physiques, alors, en vertu de la relation d'identité²¹, ils doivent avoir exactement les mêmes propriétés que ces derniers, incluant les propriétés causales²². Mais le monisme anomal n'est pas la seule version possible du physicalisme non réductionniste. Celui-ci peut s'appuyer sur la notion de constitution, ou de « composition », qui est différente de celle d'identité²³. Or, ce qui est vrai des particuliers, comme les objets, peut bien l'être également des événements²⁴. Les événements peuvent être composés d'autres événements, mais si la composition n'est pas l'identité, alors nous pouvons nier la thèse de l'identité token-token, sans rejeter pour autant la thèse physicaliste puisqu'il semble tout à fait plausible que tous les événements soient composés d'événements physiques.

Considérons l'exemple paradigmatique de particuliers concrets que sont les objets. Je soutiens que deux objets numériquement distincts, par exemple une certaine masse de matière et vous-mêmes, peuvent coïncider dans une certaine région de l'espace et du temps. Les arguments en faveur de cette approche métaphysique insistent particulièrement sur le rejet de l'essentialisme méréologique²⁵, et ils présupposent que chacun des deux objets nommés existent réellement comme des entités distinctes, malgré leur coïncidence dans l'espace et dans le temps²⁶. Évidemment, l'existence de tels objets coïncidents porte à controverse, mais je considère très sérieusement la plausibilité d'arguments comme les suivants²⁷. Considérons une petite statue, à un temps *t*, disons une réplique en argile du Discobole que vous avez placée sur une étagère, ainsi que la masse d'argile à partir de laquelle cette statue a été fabriquée. Au temps *t*, la statue et la masse d'argile ont toutes

21. Je ne suis pas un fan de la notion d'identité contingente, dont j'ai du mal à comprendre la cohérence.

22. Je parle de propriétés plutôt que de prédicats, tout comme le fait Davidson dans « Mental Events », bien que Davidson soit officiellement un nominaliste.

23. Voir Doepke, « Spatially Coinciding Objects » ; Thomson, « The Statue and the Clay » ; et Johnston, M., « Constitution is Not Identity », *Mind*, 99, 1998.

24. Les objets et les événements sont différents, mais je crois qu'ils partagent la propriété d'avoir des parties. Les événements n'existent pas dans leur totalité à des temps différents, alors que les objets se répètent à travers le temps.

25. Il y a deux versions importantes de cette thèse : un objet ne peut survivre (i) à la destruction d'une de ses parties, et (ii) s'il perd une de ses parties. Ces thèses semblent être fausses de la plupart des objets macroscopiques.

26. Peter Van Inwagen rejette l'existence de votre corps et celle d'artefacts ; voir Van Inwagen, P., *Material Beings*, Ithaca, Cornell University Press, 1990. Donc, selon lui, ces arguments sont faibles. Mais notons qu'un effort considérable d'argumentation non circulaire doit être déployé lorsqu'il s'agit de nier l'existence de notre propre corps.

27. Voir Doepke, « Spatially Coinciding Objects », et Thomson, « The Statue and the Clay ». L'exemple qui suit s'inspire d'exemples similaires de Thomson et de Baker, « Why Constitution Is Not Identity ».

leurs parties physiques en commun, mais il n'est pas nécessaire qu'elles aient toujours toutes leurs parties physiques en commun. La statue peut survivre à la perte de l'une de ses parties si, par exemple, lors d'une crise de dégoût esthétique, vous en brisez un des pieds que vous jetez ensuite par la fenêtre. Votre statue est endommagée mais elle est toujours sur l'étagère. Il n'en va pas de même pour la masse d'argile. Si celle-ci existe encore, ce qui paraît douteux, elle se sera dispersée : une de ses parties sera toujours sur l'étagère et le reste se retrouvera dans la cour. Briser l'un des pieds ne fait qu'endommager la statue, mais cela semble tout simplement détruire la masse d'argile ou, à tout le moins, la disperser dans l'espace²⁸.

Certains philosophes ne sont pas convaincus par ces arguments en raison d'un certain scepticisme relativement au statut ontologique des artefacts²⁹. Mais les arguments de ce genre semblent encore plus plausibles pour les « objets » humains :

- 1) Hermès peut perdre un pied et être toujours là où il était.
- 2) La masse de chair qui, à un certain temps *t*, coïncide dans l'espace et dans le temps avec Hermès ne peut perdre la partie qui coïncide avec le pied d'Hermès et être toujours là où elle était.

Donc : Hermès n'est pas la masse de chair.

Comme dans le cas de la statue, l'amputation du pied d'Hermès ne le détruit pas et ne le disperse pas. Mais si on considère la masse de chair dans son entier, alors ou bien l'amputation la disperse ou bien elle la détruit (et la masse de chair qui reste serait alors différente). Donc, Hermès n'est pas identique à la masse de chair. D'autres arguments de ce type peuvent être construits en utilisant des propriétés essentielles³⁰ :

- 1) Hermès est essentiellement une personne ou un être psychologique.
- 2) La masse de chair n'est pas essentiellement une personne ou un être psychologique.

Donc : Hermès n'est pas la masse de chair.

Plusieurs objections peuvent être adressées à ce type d'arguments, et leur défense exige beaucoup de travail. Mon but ici est simplement de développer ce point de vue métaphysique et d'en montrer la pertinence pour la discussion de l'argument de Kim. Si nous trouvons que ces thèses métaphysiques sont plausibles, et à mon avis elles le sont, alors le rejet du principe d'héritabilité causale devient tout à fait plausible.

28. Dans « The Statue and the Clay », Thomson suggère une autre version de cet exemple : vous remplacez la masse d'argile qui répugne à votre bon goût par une autre masse. Dans ce cas, la statue semble être complètement sur l'étagère, alors que la masse d'argile originale ne s'y trouve pas complètement.

29. Voir Van Inwagen, *Material Beings*.

30. Il s'agit ici d'une adaptation d'un argument de Baker, « Why Constitution Is Not Identity ».

La différence entre Hermès et la masse de chair, que l'on peut appeler son corps, est que Hermès est essentiellement une personne, tandis que la masse de chair pourrait survivre même si les conditions requises pour que Hermès soit une personne cessaient d'être satisfaites, par exemple la condition d'être une créature qui pense. Ainsi, les deux objets diffèrent relativement à une certaine propriété, à savoir la propriété d'être-essentiellement-une-créature-qui-pense, ou celle d'être-une-personne. Selon moi, il s'agit là d'une propriété causalement pertinente des objets et des événements qui l'exemplifient³¹. La propriété d'être-une-créature-qui-pense est causalement pertinente pour les objets qui pensent : ils peuvent raisonner et agir sur la base de leurs raisonnements. Pouvoir survivre sans penser est une différence causale qui distingue la masse de chair de la personne. (Il en va de même pour la capacité de survivre à la perte d'un tiers de votre masse corporelle.) Les différences modales entre des objets sont souvent des différences causales. Considérons les deux énoncés contrefactuels suivants, qui sont des énoncés causaux à propos de moi et de la masse de chair. Il semble y avoir des différences importantes entre eux. L'énoncé « Si je cessais de penser et que je m'effondrais sous la porte, je bloquerais le passage » est faux (à moins d'utiliser « je » de façon équivoque), car une personne ne peut survivre à la perte de la pensée et une personne qui n'existe pas ne bloque rien. Mais la masse de chair n'est pas assujettie à une telle limitation. En effet, il est plausible que l'énoncé suivant soit vrai : « Si la masse de chair cessait de penser et s'effondrait sous la porte, elle bloquerait le passage. » Ainsi certains objets distincts peuvent coïncider spatiotemporellement. Ils sont numériquement distincts, car différents ensembles d'énoncés contrefactuels s'y appliquent. Et il est plausible que ces énoncés contrefactuels révèlent des *différences et des propriétés causales*.

Illustrons cette idée en considérant le rôle causal/fonctionnel d'un objet. Le rôle causal/fonctionnel d'une propriété correspond aux généralisations nomologiques où elle figure. Pour les objets, nous pouvons spécifier le rôle causal en faisant la liste des divers énoncés contrefactuels qui s'y appliquent compte tenu de diverses circonstances actuelles et contrefactuelles. Ainsi le rôle causal de Hermès et celui de la masse de chair seront différents. Les propriétés qui rendent une personne différente de la masse de chair qui la compose (ou de son corps) sont des propriétés causalement efficaces. Être-une-créature-qui-pense ou être-essentiellement-une-créature-qui-pense sont causalement pertinentes.

Mais qu'est-ce que cela implique relativement au principe d'héritabilité causale? Notons d'abord que ce principe concerne la relation d'héritabilité de pouvoirs causaux, qui est une relation entre deux particuliers : soit

31. Cet argument peut sembler circulaire mais il ne l'est pas. Kim lui-même reconnaît l'existence de propriétés psychologiques causalement efficaces, à condition qu'elles soient identifiées à des propriétés physiques homogènes. Le point de vue que j'élabore ici ne dépend pas de la thèse de la réalisabilité multiple mais uniquement de l'efficacité causale des propriétés psychologiques.

l'occurrence d'une propriété mentale (être-une-créature-qui-pense) et celle d'une propriété physique (être-une-masse-de-chair). Mais si mes arguments sont corrects, alors il existe des différences causales entre ces deux particuliers, même s'ils coïncident dans l'espace et dans le temps. L'occurrence de la propriété être-une-personne, ou être-une-créature-qui-pense, ou même être-moi (si l'on est prêt à accepter une telle propriété), n'a pas nécessairement les mêmes pouvoirs causaux que ceux de l'occurrence coïncidente de la propriété physique. La relation d'héritabilité causale n'implique pas que ce soit nécessairement ou bien tous les pouvoirs causaux qui sont hérités, ou bien aucun. Ce qui est hérité *au temps t*, c'est, dans le meilleur des cas, les causes et les effets actuels : mais ceux-ci ne constituent pas la totalité de ce qui pourrait être hérité. De plus, les occurrences de propriétés survenantes ne sont pas des indigents : elles peuvent hériter mais elles travaillent également pour gagner leur vie, en ce sens qu'elles font aussi un travail explicatif puisqu'elles entrent dans des lois *ceteris paribus* ou dans des généralisations. Une occurrence de la propriété d'être-une-créature-qui-pense a des propriétés modales distinctes de celles d'une occurrence de la propriété d'être-une-masse-de-chair, même si cette dernière est constitutive de la première.

Mais que devrions-nous dire alors des propriétés de deuxième ordre? Kim soutient qu'il n'existe d'autres propriétés que les propriétés de premier ordre : « En quantifiant sur des propriétés, nous ne pouvons créer de nouvelles propriétés, pas plus que nous pouvons créer de nouveaux individus en quantifiant sur des individus³² ». Considérons l'exemple suivant :

Être-dormitif est une propriété de deuxième ordre relativement à un ensemble de propriétés de base *B* (être-de-l'opium, être-un-tranquilisant-à-cheval, etc.) si et seulement si être-dormitif, c'est avoir une certaine propriété *P* (par exemple être-de-l'opium) telle que *D(P)* (par exemple, faire-dormir-un-sujet (opium)), où *D* spécifie une certaine condition sur les membres de *B*.

Mais qu'est-ce à dire au juste que l'opium satisfasse une certaine condition (causale)? Cela ressemble curieusement à une quantification : tous les membres de *B* possèdent la propriété de faire-dormir-un-sujet, ou la propriété de dormitivité! Les quantificateurs apparaissent deux fois plutôt qu'une dans cette définition des propriétés de deuxième ordre. Quantifier sur des propriétés n'amène pas de nouvelles propriétés à l'existence, mais quantifier sur deux ensembles de propriétés (les propriétés de base et celles qu'elles réalisent) nous engage ontologiquement à deux ensembles de propriétés. Bien sûr, le débat porte en grande partie sur cette question : y a-t-il deux ensembles de propriétés ou un seul? Si nous prenons aux sérieux les exemples que j'ai discutés (statue/argile et personne/corps), nous avons des raisons de croire qu'il y a bien deux ensembles de propriétés. Une occurrence d'opium constitue, ou compose, une occurrence de la dormitivité, mais celles-ci ne satisfont pas les mêmes contrefactuels causaux et ne sont donc pas identiques.

32. Kim, « The Mind-Body Problem : Taking Stock After Forty Years », p. 201.

Bien que je défende le point de vue selon lequel il existe des particuliers qui ne sont pas identiques aux particuliers qui les composent, dans la mesure où les premiers auraient des propriétés causales distinctes de celles des seconds, il semble assez évident que de tels particuliers et certaines de leurs propriétés ne peuvent exister qu'à condition que d'autres particuliers et d'autres propriétés existent. (C'est pourquoi ce point de vue n'est pas exactement cartésien.) Et c'est justement là que réside l'attrait d'une conception de la nature comme étant organisée dans une hiérarchie de niveaux, ainsi que la motivation à l'appui de la notion de propriété de deuxième ordre.

Cette problématique soulève plusieurs autres questions. D'où est-ce que les particuliers tiennent leurs pouvoirs causaux sinon complètement de leurs bases constitutives? Devrait-on croire que le domaine physique n'est pas causalement clos, que certains événements physiques n'ont pas de causes physiques? Je suis plutôt enclin à dire que les particuliers obtiennent leurs propriétés causales des propriétés qu'ilsinstancient. Que ces propriétés ne soient pas uniquement des propriétés physiques de premier ordre explique pourquoi les particuliers ne sont pas identiques aux particuliers qui les composent (la masse d'argile, de chair, d'atomes).

Mais qu'en est-il de la clôture causale du domaine physique, cette thèse selon laquelle tout changement physique a une explication physique complète, ou une cause physique complète? Si nous admettons que la statue soit une entité causale distincte de la masse d'argile qui la compose, n'admettons-nous pas du coup l'existence de pouvoirs causaux non physiques? Si nous admettons que les événements mentaux sont causalement distincts des événements physiques qui les composent, n'admettons-nous pas l'existence de pouvoirs causaux non physiques? Effectivement, il semble que j'admette l'existence de tels pouvoirs causaux, mais cela n'entraîne pas la conséquence additionnelle qu'il existerait des événements physiques qui ne sont pas causés par d'autres événements physiques. Autrement dit, je ne nie pas que chaque événement physique ait une cause physique complète et une explication physique complète, mais je soutiens simplement qu'il existe certains événements physiques qui sont causés par des événements constitués par des événements physiques mais non identiques à ceux-ci. Bien que le principe de clôture de la théorie physique soit correct, c'est-à-dire que tout événement physique a une cause physique, le principe de clôture *causale* du domaine physique doit être rejeté. En d'autres termes, même si chaque événement physique a une cause physique, certains de ces événements ont également une cause mentale : ils sont aussi causés par des événements mentaux distincts de leurs constituants physiques. Cette thèse peut paraître surprenante. Malgré les apparences, j'aimerais suggérer qu'elle ne se prononce pas tant sur la nature du domaine physique, mais qu'elle montre plutôt le caractère insaisissable de la nature des relations causales³³.

33. Ce texte a été traduit de l'anglais par Paul Bernier. Je remercie Antonia LoLordo et Paul Bernier pour leurs commentaires portant sur des versions antérieures de ce texte.

Les conditions de la connaissance de soi

Joëlle Proust

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004971ar>
DOI : <https://doi.org/10.7202/004971ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Proust, J. (2000). Les conditions de la connaissance de soi. *Philosophiques*, 27 (1), 161–186. <https://doi.org/10.7202/004971ar>

Résumé de l'article

La connaissance de soi suppose que l'on puisse former des pensées vraies de la forme « je [.Psi] que P », où « [.Psi] » fait référence à une attitude propositionnelle, « P » à son contenu, et « je » au penseur de cette pensée. La question qui se pose est de savoir ce qui, dans le contenu mental occurrent [P], justifie l'auto-attribution de cette pensée. Ce problème, dit de la transition, soulève trois difficultés : celle de la préservation du contenu intentionnel entre la pensée de premier et de second ordre, celle de la reconnaissance de l'attitude ayant ce contenu intentionnel pour objet, et enfin la reconnaissance que ce qui est pensé l'est par le sujet qui pense. Le présent article se propose de montrer que la troisième difficulté résiste à une approche fondée sur l'expérience ou sur la signification cognitive de [P], et avance l'idée que la notion d'action mentale permet d'éclairer les conditions d'identité du penseur de [P] et du sujet de l'auto-attribution de l'attitude propositionnelle « [.Psi] que P ».

Les conditions de la connaissance de soi

JOËLLE PROUST

proust@poly.polytechnique.fr

CREA École polytechnique

RÉSUMÉ. — La connaissance de soi suppose que l'on puisse former des pensées vraies de la forme « je Ψ que P », où « Ψ » fait référence à une attitude propositionnelle, « P » à son contenu, et « je » au penseur de cette pensée. La question qui se pose est de savoir ce qui, dans le contenu mental *occurent* [P], justifie l'auto-attribution de cette pensée. Ce problème, dit de la transition, soulève trois difficultés : celle de la préservation du contenu intentionnel entre la pensée de premier et de second ordre, celle de la reconnaissance de l'attitude ayant ce contenu intentionnel pour objet, et enfin la reconnaissance que ce qui est pensé l'est par le sujet qui pense. Le présent article se propose de montrer que la troisième difficulté résiste à une approche fondée sur l'expérience ou sur la signification cognitive de [P], et avance l'idée que la notion d'action mentale permet d'éclairer les conditions d'identité du penseur de [P] et du sujet de l'auto-attribution de l'attitude propositionnelle « Ψ que P ».

ABSTRACT. — Self-knowledge requires the capacity to think true thoughts of the form “ I Ψ that P ”, where “ Ψ ” refers to a propositional attitude, “ P ” to a propositional content, and “ I ” to the thinker of the thought of content P. The question that this requirement raises is to know what, in the *occurrent* mental content [P], justifies the self-attribution of this thought. This problem, called the transition problem, involves three difficulties: how is intentional content stable across first- and second-order thoughts? How is the attitude with this intentional content identified by the thinker? And how is the thinker of the second-order thought able to claim truly that he himself is the thinker of the first-order thought? The present paper's aim is to show that the third difficulty cannot be solved through an examination of the experience of having [P] or on the basis of the cognitive significance of P, and suggests that an analysis of mental actions in which propositional attitudes play a causal and feedback role give a better grasp on the conditions of identity of the thinker of [P] with the thinker of the propositional attitude “ Ψ that P ”.

L'un des problèmes philosophiques importants que pose la question des relations entre la première personne et l'auto-attribution d'états mentaux concerne le rapport entre la métaphysique et l'épistémologie de la première personne. Toute forme de réductionnisme psychophysique doit s'atteler à la tâche de savoir non seulement sur quelles connaissances un sujet s'appuie pour savoir qu'il est un sujet, ou de quelles sources il tire cette conviction, et déterminer la nature du processus épistémique qui produit cette connaissance ou cette conviction. Il doit également se prononcer sur la nature ontologique du sujet : s'agit-il d'une *propriété* mentale (comme celle d'avoir conscience d'un état intentionnel, d'avoir des impressions qualitatives ou d'avoir co-conscience de plusieurs états), ou d'un *individu* capable de conscience, propriétés et indi-

vidus bien déterminés pour lesquels il convient de rechercher de quelles structures physiques ils dépendent systématiquement?

Quelle que soit la réponse apportée, on peut supposer que le philosophe naturaliste sera peu tenté de considérer l'articulation de ces deux questions — la question épistémologique de la manière dont on acquiert les manières de penser à soi comme sujet, et la question ontologique de ce qui rend vraies les pensées sur soi — comme un problème d'intégration, selon le terme de Christopher Peacocke¹. L'intégration suppose l'existence d'intuitions indépendantes qu'il s'agirait de réconcilier, comme c'est le cas par exemple pour le problème mathématique (ou théologique) de l'infini. Comment l'esprit humain ou tout autre esprit fini peut-il concevoir l'infini? L'intuition que partagent beaucoup de philosophes naturalistes à propos du sujet conscient, c'est que la métaphysique du sujet n'est pas séparable, comme le sont d'autres objets de connaissance, de la manière dont le sujet est appréhendé, que ce soit par lui-même ou par les autres sujets. Cette intuition concernant le type de lien qui unit le sujet et le processus de sa construction mentale demande évidemment à être affinée : s'agit-il d'un processus épistémique pleinement justifié, ou seulement fiable, permettant de faire référence à un sujet réel qui serait le porteur de ses propriétés mentales? Ou bien s'agit-il d'un processus illusoire, qui se développerait à la faveur d'une sélection étrangère à un critère de vérité, sans que le sujet puisse jamais former l'objet d'une référence fondée, ni se voir attribuer de manière vraie des propriétés qui lui appartiendraient objectivement?

Selon le type de jugement que l'on porte sur la genèse épistémique, diverses positions métaphysiques sur la nature des sujets pourront être distinguées. Celle du *réalisme* du sujet — entité stable ou dynamique — conçu au terme d'une genèse cognitive, et celle du *réductionnisme*, qui dissout le sujet au profit d'une propriété particulière des états mentaux sous-jacents. Une fois seulement ces thèses métaphysiques établies, sera-t-il possible de rechercher la base de survenance des entités ou propriétés subjectives reconnues par la théorie. L'objectif du présent article est, plus modestement, d'établir les conditions (métaphysiques et/ou épistémiques), si elles existent, auxquelles un sujet conscient peut s'attribuer ses propres contenus mentaux d'une manière qui soit stable dans le temps et qui résiste à la diversité des contenus mentaux et des attitudes qui les régissent².

1. L'usage subjectif du je

Pour savoir si un sujet peut atteindre la connaissance de soi, il convient d'analyser le type de situation dans laquelle la référence au sujet intervient pour ainsi

1. Peacocke, C., *Being Known*, Oxford, Clarendon Press, 1999, p. 1 et suivantes.

2. La question de savoir si les conditions de l'auto-attribution ainsi dégagées sont compatibles avec une métaphysique soit réaliste soit réductionniste du sujet ne peut être abordée dans le cadre de cet article.

dire « à l'état naissant », dans des circonstances où une expérience subjective, c'est-à-dire consciente et pourvue de propriétés qualitatives, suscite directement un jugement auto-attributif. Cet usage du « je » se distingue de l'usage référentiel ordinaire des termes singuliers dans le langage public, où le terme « je » fonctionne comme un terme singulier qui renvoie à l'auteur de l'énoncé. Une occurrence d'énoncé dont le sujet est « je » implique nécessairement l'*existence* de l'objet auquel « je » fait référence, et permet également aux interlocuteurs d'*identifier* l'objet auquel « je » fait référence, étant donné le contexte de la communication. Dans la pensée, en revanche, et en dépit de l'argument cartésien à cet effet, il n'est pas évident que de la simple perception occurrente d'un état de chose ou de la pensée que *P*, puisse être dérivée l'existence du sujet qui pense cette pensée, et encore moins que l'existence d'un tel sujet s'ensuive nécessairement. On peut en effet opposer à la déduction cartésienne que, du simple fait qu'un contenu mental soit appréhendé (cru, désiré, ressenti, etc.), on peut certes dériver qu'un événement mental se soit produit (un physicaliste ajoutera qu'un événement cérébral identique à l'événement mental se soit produit). Mais la proposition qui permet d'analyser cet événement pourrait être de la forme impersonnelle « il pense », par analogie avec « il pleut », comme Lichtenberg, repris par Wittgenstein, l'a soutenu. En d'autres termes, de la simple existence d'une pensée occurrente, il ne semble pas légitime de déduire qu'un sujet qui pense existe nécessairement, pas plus que de l'existence d'un événement physique, on ne peut dériver l'existence nécessaire d'un agent qui en serait la source intentionnelle.

Cet exemple nous donne l'occasion de prévenir un parallèle trop rapide sur lequel nous aurons l'occasion de revenir plus loin. « Il pleut » se dit dans une situation sans agent, où l'information et l'intentionnalité ne jouent aucun rôle. Le cas de la relation entre la pensée et son éventuel porteur n'a pas à être antisymétrique à celui de la pluie. Si l'on suppose que toute pensée occurrente implique nécessairement un sujet, ce n'est pas au sens où le penseur serait *l'agent* de sa pensée occurrente. La relation entre le penseur et sa pensée doit être pour le moment laissée dans un certain flou ; la littérature caractérise cette relation par le terme de « possession ».

Les philosophes qui ont réfléchi sur l'auto-attribution des états mentaux ont proposé de caractériser la situation « naissante » de l'usage en pensée du *je* de manière à isoler les cas où le sujet s'attribue une pensée qu'il est le seul à pouvoir s'attribuer. Il le fait, dans ce cas, avec une autorité qui n'appartient qu'à lui. Wittgenstein³ a ainsi opposé les usages du *je comme objet* des usages du *je comme sujet*. Dans l'usage objectif, l'usage référentiel du mot « je » est fait en vue de permettre l'identification du référent par son aspect physique, c'est-à-dire sur une base essentiellement publique. Des exemples en sont four-

3. Wittgenstein, L., *Le cahier bleu*, trad. G. Granger, Paris, Gallimard, 1958, p. 66-67. Voir le remarquable commentaire de G. Evans dans Evans, G., *The Varieties of Reference*, Oxford, Clarendon Press, 1982, p. 217 et suivantes.

nis par « j'ai grandi de douze centimètres », « j'ai une bosse sur le front », etc. Dans l'usage subjectif, comme dans « j'ai mal aux dents », ou « j'essaie de lever le bras », « je » n'a pas pour fonction première d'identifier une personne particulière. Ce qui l'indique, c'est qu'il est impossible de se tromper de personne quand on dit « j'ai mal aux dents ». Alors que l'usage objectif implique la possibilité d'une fausse reconnaissance (je croyais me voir dans le miroir, mais c'est mon jumeau que je vois), la fausse reconnaissance est absurde dans le cas de l'usage subjectif ; la possibilité de l'erreur sur la personne est *a priori* exclue. Un sujet qui s'attribue un état mental bénéficie, en d'autres termes, d'une complète « immunité à l'erreur d'identification⁴ ». Expliquer la connaissance du sujet acquise par auto-attribution d'états et d'événements mentaux suppose ainsi que l'on statue sur cette immunité.

Pour Wittgenstein, cette immunité renvoie à un fait de langage : le « je » de « je souffre » n'est pas, selon lui, un démonstratif ; il ne fait pas référence à la personne qui parle ou qui ressent au sens où cet usage du *je* aurait une valeur contrastive par rapport à d'autres référents possibles :

« Comment savez-vous que vous souffrez? — Parce que je le *sens*. » Mais « je sens » a exactement le même sens que « je souffre ». Il ne s'agit donc nullement d'une explication. Le fait cependant que, dans ma réponse, j'appuie sur le mot « sens » et non pas sur le « je » indique bien que par ce « je » je n'entends pas distinguer une certaine personne parmi d'autres⁵.

Les auto-attributions subjectives sont, pour Wittgenstein, de fausses affirmations sur soi. Elles portent non sur la personne qui dit « je », mais sur l'état ressenti. Ne portant pas d'information véritable venant s'ajouter au contenu d'expérience correspondant, elles sont comparables à des tautologies comme « ici est ici ». Rien de substantiel n'est dénoté par le terme singulier « je ». Ce qui fait la force de l'argument de Wittgenstein, c'est que toute interrogation sur l'identité paraît généralement présupposer la référence à un « je » immunisée contre l'erreur : en posant la question « est-ce moi qui Ψ », le penseur cherche non pas à s'identifier lui-même, mais à établir l'identité entre lui-même et la personne qui instancie la propriété mentale Ψ .

Quelles conclusions peut-on tirer toutefois de cette remarque? Le cas de l'auto-attribution ne diffère pas ici du cas général de la catégorisation d'objet : pour s'interroger sur l'identité d'un objet, il faut non seulement disposer d'un ensemble de concepts permettant de le catégoriser, mais aussi avoir la capacité de distinguer l'objet auquel on fait référence. De même que l'identification d'un objet suppose qu'on le discrimine avant de lui attribuer une propriété, le « je » ne peut devenir le thème d'auto-attributions que s'il est déjà distingué parmi d'autres. Appelons identification primaire la discrimination qui donne

4. L'expression est de Shoemaker ; voir Shoemaker, S., et Swinburne R., *Personal Identity*, Oxford, Blackwell, 1984.

5. Wittgenstein, *Le cahier bleu*, p. 128-129.

l'accès à la référence à soi, et identification secondaire, ou réidentification, les usages ultérieurs du « je ».

La manière de résister à la thèse de Wittgenstein sur le caractère tautologique de la référence à « je » consiste à maintenir que l'identification secondaire paraît automatiquement immunisée contre l'erreur parce que l'identification primaire est supposée acquise. L'agent qui, pour des raisons développementales ou psychopathologiques, échoue dans l'identification primaire, n'a pas de connaissance de soi en mode subjectif parce que, même s'il jouit d'expériences subjectives, il est incapable de faire référence à soi. C'est donc peut-être simplement parce qu'il se concentre sur les usages corrects — secondaires — des pensées subjectives en « je » que Wittgenstein constate que ce n'est pas sur le « je » que porte généralement le contenu de connaissance pertinent.

Il existe en outre des circonstances où l'on peut légitimement se poser le problème de l'identification secondaire : se demander si l'on est *soi-même* le porteur d'un état mental ou d'une attitude propositionnelle. Il est à cet égard frappant de remarquer que l'immunité à l'erreur d'identification ne s'applique pas aux souvenirs en première personne⁶. Comment peut-on adéquatement délibérer, s'engager dans des processus de révision de croyance, résoudre des problèmes, sans avoir la possibilité de répondre de manière véridique à des questions comme : Est-ce moi qui ai eu cette idée? Est-ce moi qui ai rêvé que *P*, ou est-ce toi qui me l'as raconté?

Dans ce type de rapport subjectif, il devient possible pour le sujet de se méprendre sur ce qui a été, à une époque antérieure, son propre contenu mental. Si la réponse à ces questions *au passé* est source de connaissance, il semble plausible de dire qu'il doit exister une proposition correspondante *au présent* qui est également source de connaissance pour le sujet qui la pense.

Si l'on souhaite maintenir contre Wittgenstein que la référence à soi-même vise bien à distinguer une personne, il semble qu'on doive soit rejeter le principe d'immunité à l'erreur d'identification — dans ce cas, la référence à soi cesse d'être tautologique, et peut donner lieu à une connaissance substantielle —, soit montrer que l'immunité à l'erreur d'identification ne vaut que dans certaines conditions, et non de manière universelle — elle ne découle pas d'un principe *a priori*, mais du fonctionnement normal d'un organisme capable de former des pensées en « je ». Les deux solutions ont en commun de nier que l'immunité à l'erreur d'identification dans l'usage subjectif du « je » procède d'un principe *a priori*. La présente stratégie consiste à adopter la seconde solution, en défendant l'idée que l'identification primaire détermine bien un champ de connaissance sur soi. L'immunité à l'erreur n'est pas nécessairement le signe d'une vacuité référentielle du constituant *je* dans ses usages non publics. Examinons maintenant de plus près la transition qui s'opère dans l'auto-attribution des contenus mentaux par un sujet.

6. Voir sur ce point Shoemaker, S., « Introspection and the Self » dans Cassam, Q., dir., *Self-Knowledge*, Oxford, Oxford University Press, 1994, p. 131.

2. Le problème de la transition

La transition entre un contenu mental et son attribution fait problème dans le cas où la référence à soi dans l'auto-attribution est de type « subjectif ». Dans ce cas, nous n'avons pas de moyen d'identification physique, par exemple le repère du corps individuel, ni de point d'appui sur une référence partagée (comme « celui qui vous parle »). Le contenu mental occurrent de la pensée ne fournit apparemment pas en lui-même d'information sur laquelle l'auto-attribution pourrait se fonder. Prenons par exemple la pensée perceptive occurrente :

(1) [Le téléphone est sur la table].

Comment le penseur peut-il, à partir de cette occurrence, former la pensée auto-attributive :

(2) [Je vois que le téléphone est sur la table]?

Répondre à cette question se heurte d'emblée à trois difficultés liées au fait que nous n'avons pas encore suffisamment précisé le cadre théorique dans lequel nous la posons. La première difficulté consiste à savoir de quoi le contenu intentionnel de (1) est constitué. S'agit-il d'un contenu portant sur des *objets*, sur une *propriété* particulière — leur relation spatiale — ou sur le *fait* que ces objets ont cette propriété? La seconde difficulté consiste à préciser la nature du contenu de la perception : s'agit-il du contenu non conceptuel qui sert de base perceptive au jugement perceptif correspondant? S'agit-il d'un contenu conceptuel articulant une propriété (ou une relation) et un (ou deux) individu(s)⁷? La troisième difficulté est d'établir de quoi nous sommes conscients quand nous avons un contenu mental comme (1) ; est-ce que les seules caractéristiques dont nous sommes conscients sont intentionnelles ou représentationnelles⁸, ou bien l'expérience a-t-elle des traits qualitatifs non représentationnels qui détermine l'« impression que cela fait » d'avoir cette expérience?

Nous ne pouvons, dans le cadre du présent article, examiner toutes les possibilités que les divers types de réponses à ces trois questions déterminent. Nous déciderons sans pouvoir justifier ce choix ici que le contenu mental de (1) consiste dans un fait, c'est-à-dire un état de chose réalisé, impliquant en l'occurrence une certaine relation spatiale, [sur], entre deux individus, [le téléphone] et [la table]. Le fait qui forme le contenu mental en question est exprimé, nous le supposerons également, dans un format représentationnel

7. Sur ces distinctions, voir Dretske, F., *Naturalizing the Mind*, Cambridge (Mass.), MIT Press, 1995, p. 27 et suivantes.

8. Sur cette distinction, voir Peacocke, C., *A Study of Concepts*, Cambridge (Mass.), MIT Press, 1992.

9. Cette position est défendue entre autres par Dretske, *Naturalizing the Mind*, et Harman, G., « The Intrinsic Quality of Experience », dans Tomberlin, J., dir., *Philosophical Perspectives*, 4, *Action Theory and Philosophy of Mind*, Northridge, Ridgeview, p. 31-52.

conceptuel, ancré dans une présentation visuelle qui en fournit le contenu non conceptuel¹⁰. En d'autres termes, le contenu mental consiste dans le jugement perceptif qui fait suite à l'extraction d'une information visuelle portant sur les relations spatiales entre la table et le téléphone.

La troisième question, en revanche, ne peut être réglée aussi rapidement. Car savoir *de quoi* nous sommes conscients quand nous avons une pensée dont le contenu est (1) est crucial pour déterminer si (1) contient déjà les éléments d'une transition vers (2). Deux possibilités extrêmes se présentent. La première possibilité consiste à explorer la nature phénoménologique du contenu de conscience. Les deux individus constituant de (1) ont des propriétés telles que la couleur ou la forme, et pour l'un des deux, occasionnellement, le timbre, la fréquence, le volume du son, autant de propriétés qui sont tenues par cette théorie essentiellement inhérentes à la conscience que j'en ai, et seulement indirectement liées à l'objet intentionnel. Dans cette perspective, je ne peux pas former le jugement perceptif que le téléphone est sur la table sans avoir recouru à la phénoménologie de mon expérience, c'est-à-dire aux propriétés de mon état de conscience lorsque je perçois le téléphone sur la table.

Dans ce cas, la transition entre le contenu intentionnel (1) et l'auto-attribution de (1) dans (2) comporte trois étapes : i) j'ai une expérience phénoménologique ; ii) je découvre à partir d'elle (par inférence) le contenu intentionnel de mon expérience ; iii) j'utilise les moyens conceptuels dont je dispose pour décrire mon rapport épistémique à ce contenu. En particulier, pour connaître le type d'expérience que j'ai dans (i), je dois savoir que la couleur ou la forme me sont données visuellement.

L'autre possibilité extrême serait que seul le contenu (1) lui-même, c'est-à-dire une relation spatiale particulière entre deux individus, forme le contenu intentionnel conscient de la pensée correspondante. Ce dont le sujet est conscient est dans ce cas non *l'expérience* de (1) ou ses propriétés, mais le contenu intentionnel de (1). Cette théorie, selon laquelle la conscience phénoménale est transparente et n'a pour contenu rien d'autre que ce qui est intentionnellement représenté par l'état mental correspondant, est défendue par les tenants de la nature strictement représentationnelle des *qualia*¹¹. Pour savoir l'effet que cela fait d'avoir l'expérience de (1), il suffit de regarder ce qui forme l'objet de cette expérience.

Dans ce cas, la transition entre le contenu intentionnel (1) et l'auto-attribution de (1) par (2) est indirecte et médiate ; c'est non pas en me fondant sur mon expérience de (1) que je peux connaître (2) puisque (1) porte sur un état de chose extérieur, et non sur un état interne. Dans la perception du fait que le téléphone est sur la table, ce que je perçois directement, c'est l'état de chose extérieur, et non pas l'expérience que j'en ai. Mais comme les

10. Sur la notion de « contenu non conceptuel », voir Peacocke, *A Study of Concepts*.

11. Voir Dretske, *Naturalizing the Mind*.

propriétés des objets extérieurs sont récurrentes, et que je sais « à quoi ressemble le téléphone », la connaissance que j'ai de mon expérience du téléphone a elle aussi toutes les apparences de l'immédiateté. Elle est pourtant *inférée*. Voyons de quelle manière. Je typifie mes expériences sur la base de leur contenu intentionnel. Mon expérience visuelle présente est celle que j'ai chaque fois que je vois le téléphone. Pour pouvoir ainsi typifier les *contenus d'expérience*, il faut avoir les concepts permettant de subsumer les individus, propriétés et relations présents dans les contenus d'expérience ; pour typifier les *attitudes propositionnelles*, il faut posséder les concepts subsumant le type d'opération mentale (ou d'attitude) qui prend pour objet le contenu considéré. Ainsi, reconnaître et réidentifier un téléphone suppose que l'on ait le concept de téléphone. Reconnaître la présence d'une expérience visuelle quand je perçois le téléphone sur la table suppose que je dispose du concept de vision.

En ce qui concerne les expériences internes, comme celle de ressentir une piqûre sur le doigt, la même explication peut être utilisée. L'expérience que j'ai porte sur l'état physique de mon doigt, dont la peau a subi l'intrusion d'un objet pointu. C'est ce fait qui est consciemment perçu et non à proprement parler la douleur ; dire que *j'éprouve une douleur* dépend conceptuellement de l'état intentionnel que j'ai en percevant un corps pointu au contact de mon doigt. Je peux éprouver une douleur sans avoir le concept de douleur. Mais je ne peux pas m'auto-attribuer une expérience occurrente de douleur sans avoir, d'une part, le concept de doigt, de piqûre, et d'autre part, celui de douleur.

3. La transition comme lien nécessaire

Même s'il y a en fait de bonnes raisons de préférer la conception représentationnaliste des propriétés phénoménales (celles qui constituent la conscience que le sujet percevant a de son environnement), il n'est pas nécessaire de trancher entre les deux versions opposées du contenu mental pour notre présent objectif, qui consiste à analyser la transition entre (1) et (2). Nous allons voir qu'en fait, dans l'une comme dans l'autre manière de présenter la question, le même problème se pose : la transition entre (1) et (2) n'est possible que si l'on effectue une pétition de principe sur la justification ontologique de cette transition, c'est-à-dire si l'on présuppose l'identité entre le penseur qui se représente (1) et le penseur qui s'auto-attribue cette pensée dans (2).

Pour faire apparaître la difficulté, commençons par généraliser la formulation de notre question : étant donné un contenu intentionnel occurrent *P* donnant lieu à une attitude propositionnelle occurrente de type Ψ comment le sujet peut-il former de manière autorisée l'auto-attribution correspondante :

(3) Maintenant [P] (de type Ψ)

(4) Maintenant je Ψ que P

3.1 *L'hypothèse de la nécessité métaphysique*

Une premier type de solution mérite d'être examiné. N'y a-t-il pas entre ces deux pensées un lien nécessaire, qui justifie que la transition vaille comme connaissance de soi? Examinons en premier lieu si ce lien est d'ordre métaphysique, c'est-à-dire ancré dans la nature même des états mentaux concernés et dans l'existence d'un rapport intrinsèque liant le premier au second ordre. Écartons d'emblée un argument ici sans pertinence. Je peux avoir le contenu mental [P] sans avoir l'attitude de type Ψ ; il n'y a pas de nécessité métaphysique à ce que [P] soit vu maintenant, plutôt que simplement imaginé ou souhaité. Mais cet argument ne vaut pas dans le cas de (3), où le type de P est fixé du simple fait que P est un état mental occurrent. Le problème du lien métaphysique entre (3) et (4) s'attache à l'existence de deux états mentaux, ou à l'existence d'un état mental et à l'existence d'un *je* auquel cet état mental conduirait nécessairement, comme dans la dérivation cartésienne qui part d'un état de conscience ayant un contenu déterminé (penser, ou imaginer, ou percevoir que P) pour conclure à l'existence d'un *je* qui pense, imagine, ou perçoit que P .

L'une des façons de défendre l'existence d'un tel lien métaphysique entre (3) et (4) pourrait consister à soutenir qu'il existe une relation constitutive entre eux. On pourrait par exemple défendre l'idée que les états mentaux d'ordre inférieur tels que (3) activent nécessairement un état mental d'ordre supérieur tel que (4) parce qu'ils en sont constitutivement inséparables. Dans cette hypothèse, soutenue entre autres par Davidson¹², les états mentaux d'ordre inférieur sont ainsi constitués en partie par les états d'ordre supérieur. On pourrait également défendre l'hypothèse converse, selon laquelle les états mentaux d'ordre supérieur sont en fait constitués en partie par les états d'ordre inférieur. Dans les deux cas de figure, la thèse métaphysique portant sur le rapport de constitution entre les états mentaux (3) et (4) fonde la thèse épistémologique selon laquelle appliquer un concept quelconque suppose que l'on ait la capacité de second ordre d'appliquer le concept de concept : le contenu Φ implique le contenu Ψ (Φ) ou réciproquement.

Avant de discuter la portée de cette thèse pour résoudre le problème de la transition, il convient de distinguer les *trois* manières dont cette dérivation peut être évaluée, et qui donnent lieu à trois types de questions. D'abord, est-ce que le contenu [P] qui forme l'objet intentionnel est correctement reconnu par le sujet comme étant l'objet de sa pensée? Dans les termes de Peacocke¹³, existe-t-il une « sensibilité inter-niveaux » à la stabilité du contenu? En second lieu, est-ce que les types d'opération ou d'attitudes mentales qui sont effectivement exercés de manière occurrente à propos de [P] sont correctement iden-

12. Davidson, D., « Animaux rationnels », trad. P. Engel, dans *Paradoxes de l'irrationalité*, Combas, L'Éclat, 1991, p. 63-75; voir en particulier p. 71 et suivantes.

13. Peacocke, *Being Known*, p. 220.

tifiés par le sujet comme étant les opérations ou attitudes qui sont les siennes à propos de [P]? Enfin et surtout, le sujet a-t-il raison de considérer qu'il peut s'auto-attribuer l'état mental en question? Le « je » qui intervient dans (4) est-il légitimé par les éléments de (3)?

Résumons ces trois types de sensibilité à l'état mental qui doivent être conjugués pour qu'une auto-attribution soit effectuée avec succès :

(a) la *préservation du contenu intentionnel* de l'état occurrent d'un niveau mental à l'autre (de l'état simplement activé à la conscience réflexive de cet état) ;

(b) la *reconnaissance de l'attitude* ou de l'opération ayant pour objet ce contenu intentionnel : le sujet doit identifier non seulement un contenu de pensée, mais aussi une attitude à l'égard de ce contenu ;

(c) la reconnaissance que ce qui est pensé l'est *par le sujet qui pense*. Il ne s'agit pas, comme l'a bien vu le théoricien de l'immunité à l'erreur d'identification, d'une erreur sur la personne ; mais plutôt de l'acceptation ou du déni possibles que moi, qui m'identifie sans difficulté, aie la propriété d'être le penseur de cette pensée.

Trois arguments de niveau différent permettent de rejeter la nécessité métaphysique de la transition. Le premier consiste à s'attaquer à la relation constitutive qui est censée s'appliquer à un contenu mental et à la saisie réflexive de ce contenu. Le cas de l'expérience d'un animal non humain, qui est capable de se représenter spatialement les objets de son environnement mais n'a pas pour autant les concepts de second ordre l'autorisant à former le contenu réflexif correspondant, paraît s'opposer à la thèse constitutive. Quoique l'on puisse supposer qu'il existe des états mentaux occurrents chez cet animal, on a des raisons indépendantes de penser qu'il ne dispose pas d'états mentaux réflexifs ni d'ailleurs du concept de concept ni d'aucun concept psychologique¹⁴.

La seconde objection consiste à récuser la nécessité métaphysique de l'amplication. Il n'est pas suffisant de penser que *P* pour dériver [Je Ψ que *P*]. Ce qui peut au plus être dérivé est [*P*, de type Ψ]. La validité de la dérivation pré-supposerait que le « je » soit également partie constitutive de l'état mental occurrent, ce qui n'est pas le cas. Cette seconde objection résiste ainsi à la *petitio* cartésienne par la *petitio* contraire. On peut évidemment estimer l'argument insuffisant. Il a toutefois le mérite d'indiquer une lacune argumentative dans l'argument constitutif.

La troisième objection consiste à invoquer un argument psychopathologique, que l'on peut aussi considérer comme une expérience de pensée, montrant qu'il n'y a pas de nécessité métaphysique à reconnaître que l'on a soi-même formé une pensée de premier ordre. Pour les besoins de l'argument, on suppose qu'il existe une personne qui a l'occurrence mentale (3), et l'on montre que cette personne n'est pas métaphysiquement contrainte de s'attribuer

14. Voir Proust, J., « Mindreading in Non-Human Animals », dans *Philosophical Topics*, à paraître.

cette pensée lors même qu'elle se pose la question de savoir qui est l'auteur de cette pensée. Rien ne s'oppose en effet à ce que la personne qui pense [P] — en supposant, encore une fois, qu'elle existe indépendamment de ses états de pensée occurrents — attribue la pensée (3) à autrui. On peut bien entendu estimer que le patient se trompe. Mais on ne peut sous-estimer l'existence et la force de cette conviction, dont la clinique de la schizophrénie offre maints exemples. Le patient souffrant d'un délire d'intrusion de la pensée nie souvent qu'une pensée occurrente soit sa propre pensée. Quoiqu'il ne se trompe vraisemblablement ni dans le contenu, ni dans le type de pensée dont il s'agit (et manifeste de ce point de vue l'autorité de la première personne sur ses propres contenus mentaux), il conteste que cette pensée occurrente soit la sienne. Il attribue généralement la pensée intrusive à l'influence d'un autre penseur qui a mystérieusement gagné le contrôle de son cerveau¹⁵. Il y a dans ce cas une dissociation très claire entre la pensée en *je* de second ordre et l'attribution à ce *je* de la pensée de premier ordre [P]. Ce qui nous intéresse ici, c'est que cette dissociation est facilement compréhensible et ne menace en rien la compréhension que nous avons de ce qu'est un état mental.

Du point de vue de la nécessité métaphysique, c'est ainsi essentiellement la troisième dimension de l'auto-attribution qui fait problème. Même si l'on admet que les fonctions mentales *occuper l'état occurrent* [P]/*occuper l'état occurrent* [Ψ que P] sont en relation de constitution partielle, rien de tel ne vaut de l'énoncé d'auto-attribution (4), [Je Ψ que P]. Le référent du constituant « je » n'est pas inclus dans la pensée de l'expérience égocentrique que fait le sujet. L'ampliation ontologique qui intervient dans (4) exclut que la transition soit ancrée métaphysiquement dans (3), même si la réciproque peut être concédée.

3.2 L'hypothèse de la nécessité épistémique

Dans la longue analyse qu'il consacre au problème de la transition¹⁶, Peacocke renonce lui aussi à l'hypothèse d'une nécessité métaphysique et suggère que ce qui peut fonder la nécessité du passage de (3) à (4), étant donné que l'attitude propositionnelle est de type Ψ , tient plutôt à une nécessité épistémique¹⁷. Dans toute expérience, perceptive ou non, le penseur est relié à l'objet de manière à acquérir une information non descriptive¹⁸. Ainsi, toute expérience, qu'elle soit visuelle, auditive, olfactive, nociceptive ou émotionnelle, implique l'usage

15. Sur cette question, voir Peacocke, *Being Known*, p. 242-245 et Campbell, J., « Le modèle de la schizophrénie de Christopher Frith », dans Grivois, H. et Proust, J., dir., *Subjectivité et conscience d'agir, approches clinique et cognitive de la psychose*, Paris, Presses Universitaires de France, 1998, p. 99-113.

16. Peacocke, *Being Known*, ch. 5 et 6.

17. *Ibid.*, p. 267 et suivantes.

18. Sur ce point, voir Perry, J., « Frege on Demonstratives », *Philosophical Review*, 86, 1977, p. 474-497.

de démonstratifs. L'idée est que les démonstratifs qui présentent l'expérience dans (1), *ce* téléphone, *cette* table, impliquent l'idée de présentation visuelle. L'usage des démonstratifs permet ainsi de restreindre les contextes de manière à rendre l'implication de (1) à (2) *valide*. Tous les contextes où (1) est vrai rendent également (2) vrai.

Pour comprendre l'argument de Peacocke, il faut rappeler que Kaplan¹⁹ a introduit à propos des pensées démonstratives une distinction importante entre le contenu d'une pensée (ce qu'il appelle « objet de la pensée ») et la signification cognitive de cet objet, c'est-à-dire son « caractère », ou encore le mode de présentation correspondant. Ainsi, le contenu mental (1) ne contient-il pas à proprement parler le mode de présentation [démonstratif visuel] ; mais il ne prend sa *signification cognitive* pour le penseur que parce que les modes de présentation des individus « téléphone » et « table » ainsi que le contexte où la présentation intervient sont donnés. L'attitude propositionnelle de façon générale est déterminée non par le seul contenu, mais par son mode de présentation, c'est-à-dire par le caractère sous lequel la proposition est appréhendée. Quoique les modes de présentation ne fassent pas partie du contenu intentionnel, ils font partie du caractère associé à ce contenu pour former la pensée complète, celle qui fait l'objet d'une attitude propositionnelle particulière.

Comme le soutient Peacocke, cette analyse a l'intérêt de mettre en lumière le rôle de précondition que joue dans l'auto-attribution la maîtrise des mêmes concepts pour former et pour s'auto-attribuer une pensée sur le monde, c'est-à-dire la condition (a) évoquée plus haut²⁰. Par cette analyse est en effet garantie la transition de la vérité de l'auto-attribution dans tous les contextes où l'énoncé de premier ordre correspondant est vrai, *une fois présupposée l'existence d'un sujet qui les énonce ou qui les pense*. Toutefois cette analyse ne nous donne pas d'explication de la transition sur les points (b) et (c). En ce qui concerne le point (b), on ne voit pas sur quoi le sujet se fonde pour identifier l'attitude propositionnelle qui est la sienne : comment peut-il subsumer sa présente expérience sous le concept de « vision » ou de « croyance » ? Ce que l'analyse nous dit, c'est que pour que l'auto-attribution soit correcte, il faut qu'elle soit rationnellement justifiée en partie par le fait que le sujet a bien cette expérience ou cette attitude. Mais sur quoi le sujet peut-il s'appuyer pour parvenir à cette auto-attribution ?

4. La transition comme processus

Une première solution consisterait à dire que le sujet doit apprendre à établir qu'une instance de propriété subjective appartient à une certaine catégorie : perception, croyance, désir, etc. Il peut établir, par exemple, qu'il s'agit d'une

19. Kaplan, D., « Demonstratives : An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstrative and Other Indexicals », dans Almog, J., Perry, J., et Wettstein, H., *Themes from Kaplan*, Oxford, Oxford University Press, 1989, p. 481-563.

20. Peacocke, *Being Known*, p. 268.

expérience visuelle parce que les *qualia* qui sont associés à la saisie de la relation spatiale [être sur] sont des couleurs, des formes, etc. Comme les expériences visuelles ont en commun de provoquer des impressions subjectives distinctives, le sujet peut rassembler ces impressions dans une représentation catégorielle unique²¹. Si, comme le soutient Goldman, il y a une impression distinctive pour toutes les attitudes, ainsi que pour le degré auquel elles s'appliquent à leur contenu, la transition entre (3) et (4) consiste dans la détection par le sujet qu'il se trouve dans l'attitude. Selon cette théorie, l'introspection n'est pas à proprement parler la perception d'états internes. Introspecter consiste plutôt à effectuer la mise en correspondance entre une représentation catégorielle mémorisée et une représentation instanciée ou (dans le cas des sensations) un état instancié.

Cette analyse peut être appliquée soit aux seuls contenus d'expérience, qui seraient alors individués de manière étroite par les états psychologiques du sujet, soit aux seules attitudes propositionnelles, soit aux deux. En choisissant l'une ou l'autre de ces alternatives — Goldman lui-même choisit la dernière — on prête le flanc aux critiques externalistes bien connues. S'il est vrai que la pensée consciente conduit rationnellement à la formation de croyances sources de connaissances, il paraît nécessaire que soient individués de manière externe le contenu des croyances *et* le processus de pensée rationnelle. Si ces deux conditions n'étaient pas remplies, on n'aurait pas la garantie que des connaissances soient produites par le processus de formation de croyance. En effet, le penseur pourrait former ces croyances sans être dans les relations adéquates avec son environnement pour que les contenus de sa conscience produisent ses croyances de manière rationnelle²².

Une seconde stratégie consisterait à dire que, chaque fois qu'un penseur a une expérience particulière, il a une disposition à former automatiquement une pensée portant sur l'*occurrence* de cette expérience. Cette disposition n'est pas accessible à la personne, et ne s'appuie donc sur aucune impression distinctive. Elle survient sur un mécanisme subpersonnel, et ainsi se passe de toute raison donnée par le sujet pour former le jugement sur l'attitude qui est la sienne. Cette théorie, due entre autres à Sydney Shoemaker, est pour cela nommée « théorie de l'absence de raison²³ ». Comme le note ailleurs Shoemaker²⁴, la disposition ne s'exprime que lorsque certaines conditions sont données, en particulier lorsque le sujet dispose des concepts de soi-même et d'état mental, et se pose la question de savoir dans quel état il se trouve. Le sujet n'aura ainsi

21. Goldman, A., « The Psychology of Folk Psychology », *Behavioral and Brain Sciences*, 1993, p. 16 et 23.

22. Peacocke, *Being Known*, p. 227.

23. Shoemaker, S., « Special Access Lies Down with Theory-Theory », *Behavioral and Brain Sciences*, 1993, p. 16 et 78-79. Voir aussi Shoemaker, S., « Self-Reference and Self-Awareness », dans Cassam, dir., *Self-Knowledge*, p. 80-93.

24. Shoemaker, S., *The First-Person Perspective and Other Essays*, Cambridge, Cambridge University Press, 1996, p. 222.

de connaissance de soi que pour autant que le mécanisme subpersonnel d'auto-attribution sera fiable. Distinguons deux manières de comprendre cette fiabilité. La première est essentiellement étroitement fonctionnelle. Le mécanisme subpersonnel envisagé est constitué par un dispositif neuronal assurant que, étant donné un certain type d'entrée, un certain type d'effet auto-identificateur ou auto-attributeur s'ensuivra ; il ne produira l'effet attendu que pour autant qu'il fonctionne normalement, ce qui dépend d'un ensemble de conditions ayant trait à la chimie cérébrale, à l'état occurrent des neurones, des synapses, etc. Dire qu'un tel mécanisme est fiable, c'est dire qu'il produira subpersonnellement une information qui pourra finalement être utilisée par le sujet sans qu'il connaisse le mécanisme auquel il doit son savoir.

Toutefois, la fiabilité peut être étendue au-delà des caractéristiques fonctionnelles, causales et donc physiques du mécanisme considéré si le sujet lui-même *est capable d'apprécier la valeur informationnelle de la sortie*. Utilisant son intelligence et ses capacités conceptuelles, le sujet peut, le cas échéant, rejeter l'impression formée subpersonnellement. En voici un exemple. Il n'est pas rare que les montagnards non entraînés à la raréfaction de l'oxygène en très haute altitude soient sujets à des hallucinations. L'un d'entre eux rapporte avoir eu, à 5000 mètres, la « vision » de petits hommes marchant au pas. « J'ai l'impression de percevoir que [P], se dit alors le sujet, mais il est impossible que je voie [P] parce que [P] est sinon impossible, du moins très improbable. Je dois halluciner. » La fiabilité causale du mécanisme a, dans ce cas, été complétée par la fiabilité de l'ensemble du système rationnel de l'auto-attribution.

Du point de vue qui nous intéresse, à savoir celui de l'existence d'un lien épistémique garantissant que la transition de (3) à (4) est source de connaissance, il faut bien constater que l'invocation d'un mécanisme subpersonnel peut certes guider de manière fiable l'identification du type de l'attitude Ψ que le sujet a la disposition de s'attribuer lorsqu'il forme la pensée de contenu [P], et ainsi garantir la connaissance de l'attitude en question. Mais évidemment, ce qui n'est nullement garanti dans cette approche, c'est que le sujet attribue cette attitude ou cette opération mentale (croire, percevoir, rêver, halluciner, etc.) *au sujet même qu'il est*. Encore une fois, il ne s'agit pas de plaider pour l'inexistence du sujet, mais de maintenir que rien, dans les analyses envisagées, n'interdit la possibilité qu'un sujet qui s'identifie correctement comme lui-même puisse nier être le penseur de cette pensée.

Toutes les analyses épistémiques envisagées échouent à satisfaire la condition (c) évoquée plus haut : elles permettent au mieux de justifier l'implication épistémiquement nécessaire, non de (3) à (4), mais de (3) à (5), c'est-à-dire à :

- (5) Maintenant est formée une attitude propositionnelle ayant le contenu *P* sous une perspective mienne.

En toute rigueur, le caractère implique le sujet qui pense cette pensée uniquement de manière épistémique, c'est-à-dire à travers la pertinence égocentrique

du contenu pensé. Le caractère n'est qu'un mode de donation de la pensée, et non un mode d'introduction d'un individu qui penserait la pensée : que le caractère soit subjectif n'implique en rien le titre qu'aurait un sujet à s'attribuer l'expérience considérée.

Une autre manière de présenter cet argument consisterait à invoquer l'existence de nombreuses illusions concernant la conscience du mien. Quoique les philosophes qui s'appuient sur la conscience pour dériver l'auto-attribution défendent généralement une conception externaliste des contenus mentaux, on peut objecter que, précisément, la conscience d'être en relation subjective avec quelque chose (qui s'exprime dans l'usage adjectival traduit plus haut par « sous une perspective mienne ») n'a pas toujours d'ancrage dans deux termes indépendants objectifs, dont l'un serait le sujet et l'autre le contenu mental. Par exemple, contemplant les occupants d'un manège dans une fête foraine, il est banal d'éprouver soi-même l'impression que l'on a lorsqu'on est en haut du « Grand Huit ». Utilisant son intelligence et ses capacités conceptuelles, on peut bloquer l'auto-attribution que l'on perçoit soi-même le vide.

S'il n'est pas possible de tirer du caractère davantage qu'un rapport adverbial avec le sujet putatif de l'expérience, rien ne garantit que ce rapport adverbial ne soit rejeté par le sujet de la pensée, utilisant ses capacités propres d'intelligence et de conceptualisation.

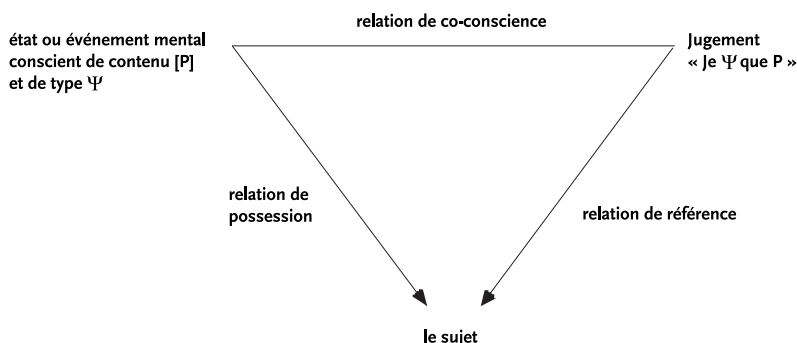
5. L'hypothèse du lien rationnel

Selon Christopher Peacocke, le contenu de l'état conscient de premier ordre ne fournit pas une *donnée* permettant d'appuyer la transition ; mais il fournit néanmoins une *raison* d'effectuer cette transition : l'auto-attribution a la propriété d'être une *transition primitive rationnellement imposée* au penseur qui a l'état conscient de premier ordre du fait qu'il dispose des concepts pertinents²⁵. Il se range ainsi parmi les théoriciens qui défendent l'existence d'un lien épistémique nécessaire entre (3) et (4)²⁶. Nous avons présenté plus haut cette position et avons vu que, lorsque nous développons les points laissés en suspens quant aux raisons que le sujet peut avoir de s'auto-attribuer ses états mentaux de premier ordre, nous ne parvenons pas à découvrir dans les termes de cette théorie la nécessité épistémique de la convergence référentielle entre une personne supposée donnée et le sujet éventuel possédant les états mentaux occurrents.

25. Peacocke, C., « Conscious Attitudes and Self-Knowledge », dans Wright, C., Smith, B. C., et Macdonald, C., dir., *Knowing Our Own Minds*, Oxford, Clarendon Press, 1998, p. 62-98. Voir l'excellente discussion par Martin, M. G. F., « An Eye Directed Outward », dans Wright, Smith, et Macdonald, dir., *Knowing Our Own Minds*, p. 99-121, en part. 111.

26. Sans exclure d'ailleurs complètement qu'il existe également un lien métaphysique entre l'existence de la conscience et l'existence d'un porteur de cet état : voir Peacocke, *Being Known*, p. 292.

Figure 1



Il peut être utile, avant de proposer une théorie de cette convergence référentielle, de rappeler dans quels termes, de l'avis de Peacocke, une théorie satisfaisante de la transition doit généralement se présenter. Il propose non une théorie substantielle de la transition, mais un schéma d'explication, c'est-à-dire une classe de théories qu'il nomme *théories-delta*, ainsi nommées parce que, prenant pour bases l'état mental de contenu [P] et de type Ψ et le jugement [Je Ψ que P], elles associent à chacune de ces bases l'unique sujet auquel le terme « je » fait référence et qui possède l'état mental de contenu [P]²⁷.

Ce schéma part de deux prémisses fondamentales, que l'on peut expliciter dans les propositions (6) et (7) :

(6) Pour tout état ou événement mental conscient, il existe un et un seul possesseur de cet état ou événement, qui est le sujet conscient.

(7) L'occurrence de l'auto-attribution d'un état mental conscient est en relation de co-conscience avec l'occurrence de l'état mental conscient correspondant.

Étant donné (6) et (7), le delta paraît pouvoir être fermé en affirmant l'identité *a priori* entre le possesseur de l'état conscient occurrent de contenu *P* et de type Ψ et la référence de « Je » dans « Je Ψ que *P* ». Étant donné (7), qui affirme le caractère co-conscient de l'état et de son auto-attribution, et par application de l'unicité du porteur affirmée dans (6), on peut apparemment déduire que l'auto-attribution, étant également l'objet d'une occurrence d'état conscient, est possédée par le sujet. Comme on va le voir, il reste à établir toutefois que ce porteur commun des *deux* états co-conscients est également l'objet auquel fait référence le *je* dans la pensée auto-attributive.

La pointe du delta donne évidemment pour ainsi dire l'emplacement de la solution recherchée : le sujet qui possède ces états est, en un sens à préciser, *a priori* identique au sujet auquel fait référence la pensée auto-attributive [Je Ψ que P]. Comment la convergence entre la référence à la première personne et le propriétaire de l'état mental est-elle assurée? En d'autres termes,

27. *Ibid.*, p. 274.

comment garantir la relation de co-conscience sans présupposer déjà l'existence d'un seul porteur pour deux états conscients différents? Peacocke invoque la démonstration fré géenne selon laquelle un état conscient implique nécessairement l'existence d'un *porteur* de cet état. Toutefois même si nous avons la certitude qu'il existe un porteur pour tout état conscient, il n'est pas par là-même garanti que ce porteur soit identique au sujet qui s'auto-attribue le contenu de conscience. Comme l'écrit Shoemaker dans un contexte voisin, « pour pouvoir m'identifier moi-même comme étant moi-même par la possession de *cette* propriété [ici : la propriété d'être le porteur de l'état conscient [P] de type Ψ], il faudrait que je sache que *je* l'observe par le sens interne, et *cette* connaissance de soi, étant le fondement de mon identification de moi comme moi-même, ne pourrait pas être elle-même fondée sur cette identification²⁸ ». Peacocke est parfaitement conscient de cette difficulté, et rejette à juste titre toute conception de la transition qui la fonderait sur une observation quelconque²⁹.

Affirmer que seule la présupposition de l'existence d'un sujet auquel il est fait référence par les pensées en *je*, sujet qui serait identique au possesseur des états mentaux conscients, permet de fonder rationnellement les auto-attributions et par là, la connaissance de soi, peut constituer un argument transcendantal si l'on dispose déjà de la preuve indépendante de l'objectivité de la connaissance de soi. Mais il n'a pas la valeur d'une preuve directe de l'unité du possesseur d'états conscients et du *je* qui s'attribue ces états.

Nous pouvons tirer les enseignements des discussions précédentes en rappelant deux thèses classiques. La première rappelle les conditions quasi indexicales requises par l'auto-attribution. Pour atteindre la connaissance de soi-même comme étant soi, il faut davantage qu'une simple identité entre un *je* et l'attribution d'une propriété particulière à ce *je*. Comme Hector-Neri Castaneda l'a montré³⁰, « les propositions sur un *je* donné ne peuvent être des objets pleins de croyance que si la croyance en question appartient au même *je* ». L'usage normal du pronom personnel « je » ne garantit pas cette commune appartenance. L'usage quasi indexical, qu'il note *je** (*il**, *nous**, etc.), marque l'identité dans le discours oblique de l'auto-attribution entre le sujet qui forme la croyance et l'objet — le sujet lui-même — sur lequel porte cette croyance. Sans cette capacité de faire référence quasi-indexicalement au moyen d'une désignation qui engage réflexivement l'usage précédent du pronom personnel, on pourrait certes acquérir des éléments d'information portant en fait sur soi, mais non avoir la connaissance réflexive correspondante³¹.

28. Shoemaker, *The First-Person Perspective and Other Essays*, p. 88.

29. Peacocke, *Being Known*, p. 273.

30. Castaneda, H.-N., « On the Phenomeno-Logic of the I », dans Cassam, dir., *Self-Knowledge*, p. 160-166.

31. Shoemaker, *The First-Person Perspective and Other Essays*, p. 86.

C'est à utiliser le quasi indexical dans le cas de la proposition (4) que nous nous efforçons jusqu'à présent sans succès.

Le second enseignement, c'est que même à supposer que l'existence d'états conscients provoque causalement ou conditionne épistémiquement la transition vers un état d'ordre supérieur, un nouvel argument doit être produit pour tirer de ce qui ne peut être qu'une notation adverbiale de cette pensée de second ordre — comme dans (5) — une relation d'appartenance à un sujet.

Dans ce qui suit, nous tenterons de montrer que la seule façon de justifier l'unicité du sujet de la pensée comme porteur de ses états mentaux consiste à examiner dans quelles conditions effectives ce sujet pense et est amené à former des pensées sur le fait qu'il pense. L'examen de ces conditions fait ressortir que le sujet pense pour agir. La thèse que nous défendrons consiste à tirer la métaphysique du sujet de la capacité même de former des états mentaux de second ordre pour réguler son action.

6. L'appropriation des états mentaux et l'agir mental

C'est la transition qui m'attribue à *moi-même* — au sens quasi indexical de Castaneda — le contenu mental *occurent* qui nous intéresse maintenant, la transition proprement appropriative (par opposition à la transition du premier au second ordre). Non plus le fondement du passage d'un état mental conscient à la pensée de cet état mental (et non d'un autre), mais à la pensée que c'est moi qui l'ai.

Les difficultés précédentes peuvent être diagnostiquées comme provenant de la manière exclusive dont le *je* a été analysé : en relation avec le seul contenu informationnel de ses propres pensées. Il ne suffit pas d'examiner les conditions où le sujet s'auto-attribue un contenu mental donné pour obtenir le fondement de son unicité. L'unicité dont nous parlons ne peut être seulement momentanée, liée à l'existence d'un instant unique pendant lequel divers contenus de pensée sont co-conscients. Car rien ne peut apparemment interdire que se constituent en parallèle autant de sujets que d'états conscients ou de paires de tels états, étant donné la dépendance entre « être le possesseur d'un état conscient » et la relation de co-conscience entre les états mentaux possédés par un seul possesseur. Rien non plus ne vient fixer les conditions temporelles d'individuation des personnes. Combien d'instants successifs une personne doit-elle occuper pour être une personne? L'une des erreurs que nous avons commises réside dans l'idée que c'est dans la seule conscience *occurrente* de ses propres états mentaux que réside le principe de l'unicité du possesseur de la pensée.

Nous avons des raisons indépendantes d'élargir les bases sur lesquelles examiner la question de l'unicité du « je » et de prendre en compte, outre ce que le sujet sait originellement de lui-même, des connaissances rapportées au sujet — dont il n'est pas lui-même la source et qui sont formulées en troisième personne, ainsi que les autres applications du concept de sujet. Ces raisons tiennent à l'exigence générale que doit remplir tout concept, à savoir *la*

*contrainte de généralité*³². Pour comprendre le concept de « je », il faut pouvoir faire varier les deux « séries de pensées » qui entrent dans la compréhension de « j'ai *F* ». La première consiste dans la série « j'ai *G* », « *x* a *G* », « *y* a *G* ». Comme l'écrit Strawson, « une condition nécessaire pour que quelqu'un s'attribue à lui-même des états de conscience et des expériences, comme il le fait, est qu'il puisse également les attribuer, ou qu'il soit disposé à les attribuer, à d'autres que lui-même³³ ». Gareth Evans ajoute à cette première condition la disposition à comprendre une deuxième série : « j'ai *F* », « j'ai *G* », « j'ai *H* », etc. Comme l'observe Evans, « nous sommes parfaitement capables de saisir des propositions nous concernant que nous sommes entièrement incapables de dire vraies ou fausses, ni même de commencer à justifier. Je peux comprendre la pensée que j'ai été nourri au sein, par exemple, ou que j'ai été malheureux le jour de mon premier anniversaire³⁴ ». L'idée qu'Evans élabore ici, c'est que pour pouvoir avoir l'idée de soi comme sujet qui s'auto-attribue une connaissance acquise sur la base de son expérience consciente, il faut pouvoir saisir l'identité entre le « je » ainsi visé et d'autres identifications par autrui de la personne à laquelle « je » fait référence, sur la base de propriétés que la personne en question n'a pas connues directement.

Les propriétés pertinentes pour l'unicité du *je* de l'auto-attribution seront alors tantôt psychologiques, tantôt physiques ; tantôt auto-attribuées, tantôt rapportées par autrui. Le fait que ces propriétés ne soient pas toutes construites sous le point de vue du sujet garantit que je puisse me considérer comme un objet du monde objectif, une personne parmi d'autres. Non pas au sens où les autres me voient et me constituent comme personne — car je peux être une personne même sans que cette reconnaissance me soit *de facto* accordée. Mais au sens où les prédicats qui sont auto-appropriés par un sujet sont des propriétés quelconques qui élèvent leur possesseur à l'objectivité (l'indépendance à l'égard de ma pensée et de celle d'autrui). Les propriétés physiques, de type spatiotemporel, comme [être né à Lausanne le 1^{er} octobre 1900] fournissent un ancrage objectif qui font de la personne un objet du monde parmi d'autres. Pas plus qu'on ne peut identifier d'autres sujets en les identifiant uniquement comme sujets d'expériences, comme possesseurs d'états de conscience, comme l'observe Strawson³⁵, on ne peut s'identifier soi-même comme unique sur la base de ses seules propriétés mentales. Pour résumer la stratégie proposée par Evans, la meilleure façon de garantir l'unicité du « je » à travers la série de ses propriétés mentales auto-attribuées consiste à traiter ces auto-

32. Sur la contrainte de généralité, voir Strawson, P. F., *Les individus*, trad. fr. par A. Shalom et Paul Drong, Paris, Le Seuil, 1973, p. 110 et suivantes. Voir aussi Evans, G., *The Varieties of Reference*, Oxford, Clarendon Press, 1982, p. 209.

33. Strawson, *Les individus*, p. 110-111.

34. Evans, *The Varieties of Reference*, p. 209.

35. Strawson, *Les individus*, p. 112.

attributions comme un cas particulier d'attributions, et de saisir le concept de « je » dans le réseau plus général que ces attributions constituent.

Même si cette stratégie atteint le but recherché, qui est de saisir le « je » comme unique à travers les manifestations de sa conscience réflexive, il n'est pas sûr qu'elle soit la seule, ni la plus économique. Si elle est suffisante, la contrainte de généralité n'est peut-être pas nécessaire pour l'unicité du « je » qui se connaît comme possesseur d'états intentionnels. On peut en effet objecter que le sujet au sens plein d'Evans, celui qui est capable de faire le lien entre les propriétés qu'il s'attribue et celles qui lui sont attribuées, représente une forme achevée de personne qui n'est pas nécessairement mise en jeu par l'unicité du Cogito. Plus profondément, on peut douter que les propriétés mentales soient une base *insuffisante* pour fonder l'identité du « je » qui pense. Car comment un tel sens de soi pourrait-il être ancré, si ce n'est en définitive dans des croyances et des motivations? Pour pouvoir revenir mieux armé sur cette possibilité, il faut procéder à un second recentrage de l'enquête, en explorant cette fois les contextes concrets où s'effectue mentalement la réidentification du « je ».

Notre première erreur était de nous être attachés exclusivement aux états occurrents. La deuxième erreur que nous avons commise consiste à n'avoir considéré que les seules *attitudes propositionnelles* occurrentes du sujet, dans la transition entre :

(3) Maintenant [P] (de type Ψ)

et

(4) Maintenant je Ψ que P.

Une théorie alternative concernant le type de propriétés mentales pertinentes pour la réidentification d'un *je* identique consiste à examiner non pas une attitude occurrente, mais le réseau dont elle fait partie dans le raisonnement pratique. Esquissons les grandes lignes de cette théorie, avant de revenir plus loin sur les arguments qui plaident pour elle. L'unicité du « je » dans (4) ne se comprend que s'il existe plusieurs attitudes propositionnelles, Ψ , X , Φ , Ξ , etc., qui font ou ont fait l'objet d'une auto-attribution, et si ces attitudes propositionnelles ont formé un réseau de raisons ou ont été conjointement exploitées dans le cadre d'une action déterminée. La mise en perspective temporelle et fonctionnelle de cette série d'états intentionnels permet de conférer un sens plus clair à l'unicité du sujet possesseur de tous ces états. Ce qui réunit tous ces usages du « je » comme sujet d'auto-attributions réside non pas véritablement dans l'observation que les diverses attitudes propositionnelles sont *causées* par le même agent, mais plus profondément dans le fait que le même individu est *concerné* par ces attitudes propositionnelles fonctionnellement distinctes, dans la mesure où elles ont dirigé son action et où elles ont impliqué un engagement rationnel. À la différence du porteur d'une pensée, dont l'individuation spatiale et temporelle est floue et dont le rôle fonctionnel peut paraître superflu, c'est-à-dire de type « homonculaire », l'agent d'une action peut être individué plus clairement par les propriétés fonctionnelles de l'action.

Cette nouvelle hypothèse sur le « je »³⁶ ne consiste pas à identifier simplement le « je » et l'« agent » de l'action. Elle s'appuie sur la vérité banale — mais essentielle — selon laquelle les diverses attitudes propositionnelles, perceptions, croyances, désirs, espoirs, qui sont réflexivement présentées dans les pensées de second ordre correspondantes, constituent les éléments déterminants d'une action individuelle d'un type particulier, que nous appellerons « l'action réfléchie ». C'est au niveau de l'action réfléchie qu'est jugée l'intégration relative des attitudes propositionnelles entretenues par l'agent, et c'est parce que les croyances et les désirs ont pour finalité une action individuelle accomplie dans un contexte particulier donné que les attitudes propositionnelles doivent être hiérarchisées et rendues cohérentes entre elles. L'unicité de la personne constitue alors le versant normatif et stable au fil du temps de l'unicité de l'agent au fil de ses actions.

Dans la présente perspective, l'agent ne joue pas véritablement de rôle causal dans l'action : ce sont les états intentionnels de l'agent qui occupent exclusivement ce type de rôle. On ne peut pas, sur la base de cette intuition, tout simplement réduire le sujet identique à soi à un sous-ensemble de ses états mentaux — à ceux qui auraient joué le rôle causal en question. Car ce sous-ensemble ne cesse de se modifier, et l'on ne disposerait dans ce cas d'aucune capacité réidentificatoire du « je ». En revanche, il est possible, dans cette analyse, de considérer que le « je » est constitué par l'engagement normatif propre à tout agent ayant une information réflexive sur ses propres états à *réviser* ses propres dispositions si une incohérence ou une inadéquation moyen-fin apparaît et, plus généralement, à répondre rationnellement de ses actions en faisant valoir le contenu de ses attitudes dans leur rapport avec l'action envisagée. Il y a ainsi un lien direct entre la capacité de former des états mentaux de second ordre, l'aptitude à réviser les modalités de son action, et l'engagement envers ses propres attitudes ainsi révisées rationnellement³⁷. Ce qui, dans cette perspective, donne les conditions nécessaires de la réidentification du sujet comme le même, réside dans la persistance d'un engagement à l'égard d'un ensemble d'attitudes propositionnelles, cet engagement se traduisant par la disposition à réviser ses attitudes le cas échéant.

Une expérience de pensée inspirée par Akeel Bilgrami³⁸ permet d'établir et de préciser le rôle de l'action dans la réidentification du « je ». Supposons un être nommé Oblomov, comme le célèbre personnage du roman d'Ivan Gont-

36. On trouvera dans le chapitre d'Akeel Bilgrami intitulé « Self Knowledge and Resentment » une analyse de la connaissance de soi assez proche de celle qui est exposée ici, dans Wright, Smith et Macdonald, dir., *Knowing Our Own Minds*, p. 207-241. Voir aussi Frankfurt, H., *The Importance of What We Care About*, Cambridge, Cambridge University Press, 1988.

37. Bilgrami, A., « Self Knowledge and Resentment », p. 235 et suivantes.

38. *Ibid.* La présente exposition de l'expérience de pensée est différente de celle de Bilgrami dans les détails, dans le développement ultérieur, autant que dans l'arrière-plan philosophique.

charov, qui aurait la propriété d'être *entièrement* passif. Même s'il a des états mentaux conscients, supposons que rien de ces états mentaux ne passe par un contrôle, un acte d'attention, une orientation délibérée, une sélection. Supposons en outre qu'il n'ait pas la faculté de faire valoir des préférences, ni de viser des objectifs particuliers. Supposons que sa pensée soit elle aussi entièrement réactive : que les pensées se produisent dans son esprit sans qu'il ait rien fait pour les produire ou les retenir. Peut-on dire qu'un tel individu puisse atteindre une connaissance de soi?

Il ne le peut pas pour une raison massive : il ne vaut pas la peine de fixer des croyances ou des désirs si l'on ne peut pas agir sur leur base. Ce n'est pas là un fait empirique. C'est un aspect conceptuel constitutif des états mentaux, qui nous oblige à réviser notre description initiale d'Oblomov. Si des croyances et des désirs pouvaient être fixés sans faire aucune différence pour le comportement individuel, ces croyances et désirs seraient dépourvus de toute valeur normative, et seraient par conséquent des indicateurs non fonctionnels. C'est en exerçant ses dispositions à agir que l'on sélectionne les désirs efficaces sur la base des croyances et des préférences que l'on a. Un sujet qui serait exclusivement mû par un mécanisme causal ou un agent extérieurs n'aurait que faire d'états intentionnels, et *a fortiori* d'un « je ».

Imaginons maintenant qu'un autre sujet, Oblatov, soit affecté d'une incapacité un peu différente de celle d'Oblomov. Au lieu d'être entièrement passif, il peut agir, mais uniquement sur la base de ses états de premier ordre. Incapable de former réflexivement des pensées de second ordre (et d'être informé du fait qu'il croit, qu'il perçoit ou qu'il désire), il ne peut pas non plus former des désirs de second ordre sur ses états intentionnels de premier ordre : désirer s'informer davantage, rétablir la cohérence dans ses préférences, faire porter ses désirs de premier ordre sur des objets dignes d'être possédés, etc.

Notons ici que ce qui fait défaut à Oblatov ce n'est pas, comme chez Oblomov, toute capacité à agir. C'est seulement la capacité à agir d'une manière réfléchie³⁹. Oblatov est un agent, et utilise donc ses croyances et ses désirs pour agir. Mais Oblatov manque d'une capacité cruciale : celle de pouvoir rendre raison de ses choix. Il ne peut pas le faire parce qu'il n'a pas accès au contenu de ses propres états : il n'a pas de croyances sur ses désirs de premier ordre. Si un tel sujet avait des croyances incohérentes et qu'il eût conscience de cette incohérence, il devrait penser que « ce sont les *faits* qui sont incohérents », comme le note Shoemaker dans un autre contexte⁴⁰.

Il est intéressant de noter qu'Oblatov n'est pas pour autant entièrement dépourvu de toute capacité de délibérer, si l'on entend par là la capacité de faire émerger le plus puissant de ses désirs. Le mode de délibération qui lui

39. Oblatov a une parenté avec l'individu « irréflexif » décrit par Frankfurt. Mais alors que ce dernier est irréflexif par « indifférence » de sa volonté, Oblatov n'a pas la possibilité de s'y intéresser.

40. Shoemaker, *The First-Person Perspective and Other Essays*, p. 29.

est fermé consiste seulement dans la délibération conçue comme « entreprise autocritique⁴¹ », impliquant la reconnaissance de l'éventualité que certaines croyances soient fausses, et de la nécessité de procéder à de nouveaux tests.

La réflexion sur ce cas permet de dissocier clairement la capacité d'agir de l'existence d'un « je ». Même si l'on admet qu'Oblatov est un agent, il n'est évidemment pas une personne, un sujet réidentifiable, parce qu'il est constitutif d'une personne d'avoir la disposition à rendre raison de ses actions. Oblatov ne peut pas en rendre raison parce qu'il n'a pas accès à ses propres états mentaux. Remarquons que, dans cette hypothèse, la personne n'est pas une *condition* de l'action — les attitudes propositionnelles appropriées le sont, ainsi que l'intégrité du système perceptivo-moteur — mais une *conséquence* de la capacité, développée à la *faveur* de l'action, de rendre raison de son action sur la base du contenu de ses attitudes propositionnelles. L'unicité de la personne est le produit du fait que l'action se trouve *de facto* à ne concerner qu'un seul agent, celui qui a mis en jeu ses propres attitudes propositionnelles pour réguler son propre comportement. Il s'ensuit une conséquence que l'on peut juger étrange. Si, dans une société anti-individualiste, plusieurs agents répondaient régulièrement de leurs actions — exclusivement collectives — en invoquant des attitudes propositionnelles distribuées entre eux, il faudrait dire qu'il s'agit d'une seule personne. Ce qui, dans cette conséquence, est difficile à admettre, consiste non dans l'identification d'une « personne morale » — conséquence théoriquement intéressante — mais dans l'absence de rapport entre raisons données et comportement subséquent. Les conditions de la réidentification de la personne sont alors entièrement tournées vers le passé, et le rapport entre action et normativité est alors perdu.

Une troisième direction d'enquête devrait nous permettre de résoudre cette difficulté. Il s'agit cette fois d'exploiter le fait que les attitudes propositionnelles, lorsqu'elles sont révisées réflexivement, doivent faire l'objet d'un contrôle, suivi des transformations rationnelles *correspondantes* (visant à supprimer les contradictions, réviser les préférences, etc.). Or ce contrôle qui permet à l'individu de réajuster ses états mentaux aux normes de la rationalité, correctement compris, doit nous offrir une prise nouvelle sur la question de la réidentification de soi. L'idée nouvelle sur laquelle s'appuie cette précision de l'hypothèse consiste à dire qu'un sujet se réidentifie à travers *les actions mentales* qu'il effectue, où il est à la fois *évaluateur* de ses attitudes propositionnelles, *agent* des modifications qu'il aura globalement estimées nécessaires, et *support* des propriétés auto-affectées. Nous avons maintenant une triple condition qui garantit la possibilité d'une réidentification par la reconnaissance de la coïncidence (*dans la même action mentale*) entre l'agir et le pâtir, c'est-à-dire d'une autotransformation. Expliquons-nous.

Une action mentale peut être définie comme un type d'action qui, comme toute action, est causé par des désirs et des croyances, et dont le contenu inten-

41. *Ibid.*, p. 28.

tionnel est d'obtenir une propriété mentale nouvelle en utilisant à cet effet des moyens qui sont eux-mêmes mentaux⁴². Ainsi défini, ce concept d'action mentale ne s'applique généralement pas aux opérations de premier ordre telles que percevoir *P*, croire *Q*, désirer *R*. Car ces opérations de premier ordre n'ont généralement fait l'objet d'aucun contrôle ; elles n'ont pas été effectuées en vue d'obtenir une certaine propriété mentale. Elles se sont produites dans l'esprit du sujet sous l'influence de causes variées, endogènes ou exogènes. En revanche, l'attention contrôlée constitue une action puisqu'elle se développe sur la base de la croyance qu'il existe une propriété mentale qui mérite d'être atteinte (une nouvelle représentation perçue, que seule l'attention permet de former). Percevoir attentivement que *P* se distingue ainsi fondamentalement de percevoir que *P*. De même, le souvenir contrôlé, par opposition au souvenir automatique, constitue un acte mental. Chercher à se souvenir d'une date se distingue du rappel automatique d'une date. Modifier délibérément ses désirs ou ses préférences à la lumière de ses connaissances et de ses valeurs globales est une variété d'action mentale, en vue de rendre efficaces de nouveaux désirs, ou de nouvelles échelles de préférences. Comme l'a soutenu Harry Frankfurt, une différence capitale pour le concept de responsabilité oppose les désirs de premier et de second ordres : les désirs de second ordre sont indispensables pour contrôler l'opération des désirs de premier ordre.

Revenons au problème de la réidentification munis du concept d'action mentale. Il semble possible d'assembler enfin le porteur d'un état conscient et le sujet qui s'auto-attribue cet état conscient en invoquant le fait que le premier est la cible du second, tandis que le premier est le témoin et le guide de cette transformation : c'est le porteur d'un état conscient que l'action mentale cherche à modifier ; ce porteur sera directement *affecté* par la modification : s'il s'agit d'attention, il percevra un état de choses qu'il n'aurait pu percevoir sans cette action. S'il s'agit de désir, il désirera ce qu'il ne désire pas encore (ou cessera de désirer ce qu'il désire). En outre, l'action mentale s'appuie, comme toute action, sur un feedback, sur un retour informationnel qui permet d'évaluer si l'action a été ou non réussie : il est incohérent de continuer à faire attention si la propriété perceptive a été saisie, ou de désirer si la propriété motivationnelle a été atteinte. L'auto-attribution ne peut donc s'arrêter à une simple et unique attribution. Elle doit ressaisir l'auto-attribution dans le cadre de l'action mentale dont elle constitue un moyen. La réflexivité des états mentaux reflète la propriété de l'action de prélever l'assurance que les conditions de satisfaction de l'action mentale sont réunies, ou en voie de l'être. La référence au sujet dans la transition entre une simple opération mentale et son auto-attribution est indissociablement une référence au sujet qui est le porteur d'un état conscient, qui évalue ses états et les modifie parce qu'il est concerné par cette modification.

42. Pour une analyse plus détaillée, voir Proust, J., « A Plea for Mental Acts », *Rapport du CREA*, 1998.

Si cette analyse est correcte, nous pouvons la vérifier rapidement en l'appliquant au cas de notre expérience de pensée. Supposons maintenant un personnage, Oberov, capable d'agir physiquement, capable de former des états de second ordre de manière automatique, mais incapable d'agir mentalement. Cet individu ne peut pas viser l'obtention d'une propriété mentale quelconque : il ne peut pas faire attention, souhaiter connaître, chercher à se rappeler, espérer voir, craindre de s'ennuyer, etc. À la différence d'Oblomov, et à l'instar d'Oblatov, Oberov peut agir. À la différence d'Oblatov, des pensées de second ordre peuvent se former en lui. Il ne peut pas plus tirer parti de ces pensées de second ordre qu'Oblomov de ses pensées de premier ordre, mais on doit lui reconnaître une intentionnalité du fait qu'il utilise ses pensées de premier ordre pour guider son action. Oberov peut-il dire « je » en faisant référence par là au porteur des états conscients qui sont les siens quand il a des pensées de second ordre? Dans la mesure où Oberov manque des moyens cognitifs de s'auto-affecter mentalement, rien ne garantit l'unification de ses diverses auto-attributions. Oberov, proche des primates non humains, ne peut pas former de concepts psychologiques. Il ne peut pas reconnaître la causalité du mental, ni l'efficacité des croyances dans le comportement. Son concept de « je » sera dépourvu de la normativité inhérente à la capacité de réviser ses jugements : il ne fera pas référence à une personne stable à travers les contextes de l'action.

On pourrait ici avancer que la solution proposée présente un inconvénient majeur, dans la mesure où elle interpose entre le penseur et le contenu de ses états mentaux de premier ordre, individués de manière externe, un niveau opaque constitué par les états mentaux occurrents de second ordre et leur gestion active. Ne perd-on pas dès lors le principe même de la rationalité de l'acquisition des croyances? Il y a plusieurs façons de répondre à cette objection. La première consiste à observer que l'identification du sujet est *postérieure* à la dérivation d'un état de second ordre relativement à l'état de premier ordre correspondant. Ainsi, même si la construction du sujet faisait intervenir des considérations de type internaliste, elle ne compromettrait pas le caractère externaliste des contenus mentaux de premier et de second ordres. Toutefois, on peut objecter que cette réponse n'est pas convaincante, car même si l'identification du sujet est logiquement dépendante de la capacité à effectuer la transition entre des états mentaux de premier et de second ordres, cette identification n'en commande pas moins la portée de la connaissance de soi. Faudrait-il admettre que la connaissance de soi soit d'une variété fondamentalement différente de la connaissance du monde extérieur?

La seconde manière, beaucoup plus prometteuse, consiste à soutenir que le « je », justement parce qu'il est l'entité concernée par le succès des actions mentales et l'enjeu de la stratégie dans laquelle les attitudes propositionnelles sont inscrites, doit être individué en partie de manière externaliste. Ce qui, dans la théorie proposée, fait d'un « je » un « je » n'est pas qu'il se reconnaisse comme tel ou qu'il ait accès à des *qualia* particuliers ni même à des faits concernant la vie mentale. Ce statut dépend de l'imbrication fonctionnelle entre

des attitudes de second ordre qui sont toujours elles-mêmes individuées de manière externaliste. Par exemple, dans le souvenir dirigé, l'agent constate qu'il ne peut pas spontanément se souvenir de *P*. Il se demande alors s'il *peut* s'en souvenir de manière contrôlée, et il s'engage dans le processus de remémoration de *P* en traitant toutes les inférences dont il dispose qui impliquent *P*. Rien dans ce processus ne dépend de façon cruciale de l'impression de savoir, considérée indépendamment de la question de l'impression de savoir *que P*. Les *qualia* métacognitifs, si l'on peut parler ainsi des phénomènes comme « avoir quelque chose sur le bout de la langue » ou l'impression de familiarité, ne se différencient pas des *qualia* cognitifs : ils renvoient à des propriétés du monde (en l'occurrence, des processus cérébraux). L'information cruciale qu'apporte la métacognition concerne précisément le degré de fiabilité du système cognitif ou motivationnel relativement à un objectif de connaissance ou d'action. On ne peut rejeter cette source métacognitive d'information sur la fiabilité et la disponibilité des processus de traitement sans mettre en péril tout le processus de construction rationnelle de nos connaissances du monde extérieur, dont d'ailleurs le cerveau fait lui-même partie⁴³.

43. Je remercie Paul Bernier pour ses remarques critiques portant sur une version antérieure de ce texte.

Martin Montminy, *Les fondements empiriques de la signification*, Bellarmin/Vrin, Montréal/Paris, 1998, 234 p.

Éric Grillo

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004930ar>
DOI : <https://doi.org/10.7202/004930ar>

[Aller au sommaire du numéro](#)

Éditeur(s)
Société de philosophie du Québec

ISSN
0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Grillo, É. (2000). Martin Montminy, *Les fondements empiriques de la signification*, Bellarmin/Vrin, Montréal/Paris, 1998, 234 p. *Philosophiques*, 27 (1), 187–201. <https://doi.org/10.7202/004930ar>

Études critiques

Martin Montminy, *Les fondements empiriques de la signification*

Bellarmin/Vrin, Montréal/Paris, 1998, 234 p.

ÉRIC GRILLO

Université de Paris III

Sans verser dans l'éloge convenu, et sans méconnaître certains points de divergence sur lesquels je reviendrai ultérieurement, il convient d'abord, c'est justice, de signaler les éminentes qualités de ce travail, qui font de ce texte un modèle d'analyse philosophique telle qu'elle se pratique aujourd'hui dans le contexte de la philosophie analytique nord-américaine notamment. On louera d'abord la grande clarté du propos, dont on ne perd jamais de vue la progression, scandée par les conclusions partielles des différents chapitres et résumée très fermement dans la conclusion de l'ouvrage. On louera ensuite la précision et la rigueur des analyses, l'auteur ayant toujours grand soin de restituer les thèses qu'il discute dans leur contexte d'émergence, et d'en rappeler les différentes formulations, avant d'en entreprendre l'examen critique. Enfin, on louera la pertinence des remarques et objections de l'auteur, qui contribuent à une meilleure compréhension des positions qu'il examine, et cela, qu'il les adopte ou qu'il les récuse. Là se trouve à mon sens le principal mérite de l'ouvrage. Le lecteur, qu'il adopte ou non la perspective empiriste défendue par l'auteur, tirera profit de cette lecture : le partisan, parce que sa compréhension des thèses centrales d'une conception empiriste de la signification en sera affinée ; l'adversaire, parce qu'il aura acquis la certitude qu'on ne saurait écarter d'un revers de main la position empiriste en sémantique, car de là viennent les questions et objections les plus radicales que toute théorie sémantique aura à affronter, et la conviction que la controverse, à tout prendre, est toujours plus féconde que la superbe ignorance.

Dans un texte récent, Paul Gochet écrivait à propos de Quine : « Son système nous paraît être la forme la plus robuste d'empirisme existant à ce jour. Nous ne voyons pas de critique interne à lui adresser¹ ». La lecture du texte de M. Montminy est de nature à renforcer cette conviction. Les quelques critiques « internes » que, pour son compte, il adresse à Quine, notamment sur la question du « contenu empirique » des énoncés occasionnels d'observation, ou encore sur la nécessité de préférer une théorie *distale* à une théorie *proximale* de la signification, ne remettent nullement en question les conclusions quiniennes, elles les renforceraient plutôt. J'ajouterais qu'à mon

1. Gochet, Paul, « L'empirisme relatif de Quine », dans Meyer, Michel, dir., *La philosophie anglo-saxonne*, Paris, PUF, 1994, p. 318-347.

sens, la remarque de Gochet s'appliquerait aussi bien au travail de M. Montminy : si on lui accorde la perspective qu'il adopte, et que je qualifierai pour la commodité de la discussion ultérieure un « empirisme strict », on ne peut que lui accorder ses conclusions.

Ma réserve portera plutôt sur le caractère de généralité qu'il donne à ses conclusions, généralité que je conteste, car j'entretiens pour ma part quelques réserves concernant l'adoption d'un « empirisme strict » en sémantique. À titre provisionnel, la position que je défendrai pour mon compte sera une position moyenne : un empirisme modéré, dans le cadre d'une pragmatique interactionnelle au sens de Francis Jacques². Tout en accordant à M. Montminy ses réserves face aux conceptions « traditionnelles » de la signification, tout en accordant avec lui que les verdicts que les locuteurs compétents donnent aux phrases sont les guides les plus fiables pour la construction d'une théorie sémantique qui ne veut pas rompre tout lien avec les « observables », il ne me semble pas, cependant, que l'on puisse fonder ces verdicts sur la seule composante *propositionnelle* des énonciations. Mais avant de développer ce point, il importe, me semble-t-il, de restituer dans sa linéarité et sa progression (encore une fois exemplaires) la position de l'auteur.

1.2 Reconstruction de l'argument

Comme je l'indiquais en commençant, la conclusion de l'ouvrage (p. 203-207) récapitule la démarche dans son ensemble et en fixe les principaux acquis. Le propos principal de l'auteur (p. 203) était de défendre l'idée qu'il n'est pas possible d'établir de manière satisfaisante une distinction analytique/synthétique stricte. Distinction cruciale, car si on en disposait, cela voudrait dire qu'on disposerait d'un concept de signification parfaitement déterminé et délimité, et qu'on pourrait, pour un énoncé quelconque, distinguer les contributions factuelle, sémantique et doxastique dans ce qu'on nomme sa « signification ». Or, aucune des théories envisagées par l'auteur ne satisfait cette exigence.

La démarche est conduite en deux temps. Le premier moment (chapitre 1, p. 17-53) est une attaque en règle du point de vue qui accorde un caractère *fondationnel* à la distinction analytique/synthétique et qui l'appuie sur un concept *normatif* de signification, conçue comme règle ou norme gouvernant l'usage de nos phrases. C'est la conception carnapienne qui est ici plus particulièrement visée. Cette conception échoue, nous dit l'auteur, car il n'y a pas d'explication satisfaisante de la signification qui permettrait d'identifier, pour une langue quelconque, la signification des mots, expressions et phrases, et d'en justifier l'usage.

2. Voir notamment : *Dialogiques I, Recherches logiques sur le dialogue*, Paris, PUF, 1979 ; *L'espace logique de l'interlocution*, Paris, PUF, 1985 ; « De la signifiance », *Revue de métaphysique et de morale*, 2, 1987, p. 179-218.

Le second moment de la démonstration (chapitre 2, p. 55-95) s'attaque au point de vue plus « modeste » qui entend proposer un *compte rendu descriptif* de la signification, et s'appuyer sur les pratiques linguistiques pour décrire la synonymie des phrases, et de là tracer une distinction analytique/synthétique. Mais l'observation des pratiques linguistiques ne permet pas non plus d'établir cette distinction, comme l'atteste notamment l'échec du vérificationnisme (dans la version qu'en donne Dummett, p. 63 et suivantes), qui achoppe face à l'argument du *holisme* (sémantique et épistémologique) : même en assimilant la signification au « contenu empirique » d'un énoncé, on ne peut déterminer et spécifier un tel contenu pour un énoncé isolé.

Mais, note l'auteur, pour bien comprendre la portée de cet argument, il faut préciser à la fois le contenu et la portée de la thèse du holisme, ce qui nécessite (chapitre 3, p. 97-152) un détour par la thèse de la *traduction radicale* (TR). Cette thèse met en exergue le fait que toute contrainte ou hypothèse concernant la signification doit être *rendue explicite* et mise en relation avec ce qui est *observable*, c'est-à-dire le *comportement verbal* des locuteurs compétents de la langue (p. 204). Mais, note l'auteur, Quine se montre ici trop radical : son approche *proximale* de la signification souffre paradoxalement des mêmes défauts que certaines approches traditionnelles dont il entend se démarquer : tout comme les « idées » des théories classiques par exemple, les stimulations peuvent varier arbitrairement sans perturber la communication, pourvu que le comportement verbal reste le même en présence de stimuli *distaux* (p. 110 et suivantes). Il faut donc se ranger, conclut l'auteur, du côté d'une théorie distale à la Davidson, basée sur la relation causale entre la *situation* et les réactions à celle-ci.

L'examen de la thèse de la TR permet ainsi à M. Montminy de préciser et de nuancer parfois les conclusions quiniennes. Concernant le holisme d'abord : on ne peut associer un ensemble déterminé de stimulations (ou de circonstances observables) à l'acceptation/refus d'une phrase non observationnelle. Car on ne saurait démêler (on n'a pas de procédure le permettant) dans ces phrases, ce qui relève de la *signification* et ce qui relève de la *connaissance collatérale* du monde. Concernant les phrases d'observation ensuite : selon Quine, elles constituent une catégorie à part, en ceci qu'elles auraient une signification-stimulus (p. 104 et suivantes) déterminée, et il serait donc possible de les traduire par des phrases stimulus-synonymes (p. 139 et suivantes) *indépendamment* des autres phrases de la langue. Mais, remarque M. Montminy, cela est faux pour deux raisons : d'abord, il est incorrect d'appliquer le Principe de Charité (présupposé par la TR) de façon atomique dans le cas des phrases d'observation (p. 205-206) ; ensuite parce que la signification-stimulus ne détermine pas *seule* la signification, puisqu'il faut admettre qu'une *information collatérale connue de tous* joue un rôle dans l'acceptation ou le refus des phrases d'observation. La traduction d'une phrase ne saurait donc être fixée que moyennant un choix d'*hypothèses analytiques* qui impo-

sent simultanément une traduction aux énoncés catégoriques d'observation et aux autres phrases ayant des mots en commun avec la phrase en question.

Ce qui permet une formulation plus précise de la thèse du holisme sémantique : l'usage d'une phrase individuelle ne permet pas à lui seul d'en révéler la signification ; l'erreur de Carnap et Dummett est de penser que la signification peut être spécifiée à l'aide de phrases relevant d'une certaine catégorie, celles qui expliquent « dans quelles circonstances » la phrase serait acceptée ou refusée, ou celles qui donnent « les raisons » pour lesquelles la phrase serait acceptée/refusée. L'erreur de Quine est de penser que certaines phrases ont une signification indépendamment des autres (p. 206).

Enfin, il apparaît à l'analyse que le holisme se distingue de la thèse d'*indétermination de la traduction* (IT) (chapitre 4, p. 153-202). Cette dernière, comme indétermination de la *signification* (IS, p. 169 et suivantes), est liée au rejet de la distinction analytique/synthétique. Or, le holisme sémantique est compatible avec une version de la distinction analytique/synthétique, à l'œuvre notamment dans le *moléclarisme* de Dummett. Dans cette vue, en effet, bien que la signification de chaque phrase dépende de celle d'autres phrases, on peut néanmoins assigner à chacune une signification déterminée, et établir par là une distinction analytique/synthétique. La question est de savoir comment, et M. Montminy montre qu'en fait on a peu de chances d'y parvenir, car d'après la TR, l'observation de *liens causaux* entre *verdicts* et *circonstances observables* ne permet pas de déterminer si les verdicts sont attribués *seulement* en vertu de la signification, ou font intervenir également certaines croyances, en des régularités nomologiques notamment. Quant à l'indétermination de la *vérité* (chapitre 4, p. 175 et suivantes) autre forme de l'IT, elle est liée au fait que les contraintes de la traduction radicale sont des contraintes d'*optimisation*, mais qu'il n'y a pas de principe supérieur permettant d'arbitrer différentes applications optimales de celles-ci. Reste, conclut M. Montminy, que l'IT n'est pas alarmante, car elle *ne remet pas en cause la communication et la compréhension*. Elle vient seulement de ce que les différences entre traductions rivales ne correspondent à aucune différence en principe observable dans le comportement des locuteurs. Mais elle a une conséquence *épistémologique* : concernant la théorie de la connaissance, la sémantique et la philosophie de l'esprit, « nous devons nous départir d'une notion déterminée de signification quand nous abordons les questions fondamentales touchant ces domaines » (p. 207).

2. Discussion

Selon l'angle sous lequel on la considère, la conclusion de l'auteur peut apparaître radicale ou modérée, tout en restant légitime du point de vue qui a été le sien tout au long de sa démonstration. Si la seule justification épistémologiquement acceptable des notions concernées est celle qu'on peut produire dans le cadre d'un empirisme strict, alors l'auteur a bien montré que les

notions de signification, d'analyticité, de synonymie, parce qu'elles sont dénuées de fondement empirique, sont dispensables dans le champ de la théorie de la connaissance, de la sémantique et de la philosophie de l'esprit, puisqu'elles n'apportent au mieux que des distinctions mal fondées ou superflues. C'est là la lecture *radicale* de la conclusion. Lecture quelque peu tempérée cependant par la remarque additionnelle selon laquelle ce constat d'échec n'implique nullement qu'on ne soit pas en mesure de produire un compte rendu acceptable de la communication et de la compréhension linguistique. On n'a donc pas écarté *toute* notion de signification, mais seulement celles qui, parce qu'elles sont insuffisamment fondées empiriquement, reviennent à donner de la signification un concept « transcendant », c'est-à-dire définissable *indépendamment* des pratiques linguistiques des locuteurs compétents, et susceptible de valoir pour une langue quelconque (p. 77).

Dans ce qui suit, je discuterai plus particulièrement les points suivants : ma première tâche sera de clarifier et de préciser ce que j'appelle « l'empirisme strict » de l'auteur, ce qui me permettra d'expliquer pourquoi la conception de la signification retenue comme « empiriquement acceptable » peut apparaître par trop restrictive, et pourquoi, en particulier, on peut douter que sur la seule base d'un tel concept, on puisse donner un compte rendu adéquat de la communication et de la compréhension linguistique. Je défendrai alors l'idée que la critique empiriste du concept de signification incite moins à l'abandonner qu'à le *redéfinir*, et que, dans le cadre d'une pragmatique interlocutive, on peut le faire dans une perspective qui, tout en restant *dynamique*, ne fait pas intervenir un concept *transcendant* de signification (p. 77), mais un concept qui peut être rendu *compatible* avec les critères d'acceptabilité empirique défendus par l'auteur (p. 144). J'en tirerai pour finir quelques enseignements touchant le problème de la traduction.

2.1 À propos de « l'empirisme strict »

On peut accorder la version de l'empirisme que propose l'auteur dans l'avant-propos de l'ouvrage :

Je compte me concentrer sur la légitimité empirique de ce concept [celui de signification]. Une des thèses centrales de ce livre est qu'il est essentiel de pouvoir lier toute théorie sémantique à des données empiriques en principe accessibles, qui soient susceptibles d'étayer celle-ci (p. 9).

Il s'agit là, me semble-t-il, d'un souci peu contestable. Attendre d'une théorie (fut-elle sémantique) qu'elle délimite son domaine d'objectivation et se mette en mesure d'y déterminer ses observables est assurément de bonne méthode. Mais on notera qu'il n'est encore question, à ce stade, que de *lier* la théorie à des données empiriques, sans que soit précisée la nature des liens en question. C'est là ce que je nommerai la version « faible » ou « modérée » de l'empirisme, qui, encore une fois, m'apparaît peu contestable.

Mais l'auteur n'en reste pas là, et sa thèse se *radicalise* lorsque, du constat que les seules données qu'en bonne méthode empiriste on accepte de retenir (le comportement verbal associé à des stimuli proximaux ou distaux) sont *impropres* à fonder les concepts sémantiques centraux, on conclut à l'obligation de « nous départir d'une notion déterminée de signification quand nous abordons les questions fondamentales touchant la théorie de la connaissance, la sémantique et la théorie de l'esprit » (p. 207). Ici manifestement, il ne s'agit plus seulement de *lier* la théorie sémantique à des données empiriques observables, mais bien plutôt de *dériver* les concepts sémantiques fondamentaux des *seules* données empiriques retenues.

C'est là ce que j'appelle « l'empirisme strict », et que l'auteur y souscrive (en dépit des distances qu'il prend vis-à-vis du « radicalisme » qu'ien) me paraît attesté par le fait que l'on passe, dans son ouvrage, de l'affirmation modérée de l'avant-propos à l'affirmation plus radicale de la conclusion par un réseau argumentatif serré, qu'on peut reconstruire à travers la démarche de l'auteur. C'est ce réseau argumentatif qui permet de qualifier « d'empirisme strict » la position de M. Montminy, et je voudrais le reconstruire brièvement, pour mettre en évidence : (i) les décisions méthodologiques qu'il présume ou entraîne ; (ii) les difficultés qu'il soulève ; (iii) dans quelle mesure on peut le taxer de « réductionniste ».

Je voudrais tirer argument de ces considérations pour établir que l'échec constaté (et fortement argumenté) par l'auteur autorise en fait non pas une mais *deux* conclusions. La première est celle même de l'auteur : si, même dans le cadre d'un empirisme strict, les notions sémantiques centrales se révèlent insuffisamment fondées, on doit s'en départir et renoncer à un concept strict de signification. Je voudrais pour ma part arguer que l'échec constaté autorise également une conclusion plus « libérale », qui consisterait à dire que les contraintes drastiques qu'impose la perspective empiriste *n'offrent pas une assise suffisante* pour rendre compte de ce que l'on veut traiter, en particulier de la *communication* et de la *compréhension linguistique*, et qu'on peut, dans le cadre élargi d'une pragmatique interlocutive, formuler un concept opératoire de signification qui reste empiriquement acceptable.

2.2 Le « nexus rationis » de « l'empirisme strict »

À lire de près l'argumentation de M. Montminy, on s'aperçoit que l'affirmation modérée de l'avant-propos aboutit à l'affirmation radicale de la conclusion, moyennant une chaîne de raisons qui articule cinq affirmations centrales (ci-après désignées A1 ... A5), solidaires le plus souvent de décisions méthodologiques fortes.

La première (A1) est *l'idée même d'un compte-rendu empirique* de la signification, conçu comme seule alternative face aux échecs des conceptions ayant recours (notamment) à un concept *normatif* de signification (p. 14, 55, 97). L'idée centrale ici, c'est qu'au lieu de tenter de *fonder* les pratiques

linguistiques légitimes sur un concept *normatif* (toujours problématique) de signification, il vaut mieux tenter de rendre compte de la signification à *partir des pratiques effectives* des locuteurs. L'approche, dans la mesure où elle *s'appuie* sur les pratiques, peut alors être qualifiée « d'internaliste » (notamment p. 97 et suivantes) : la signification serait *immanente* à ces pratiques, dont elle procéderait intégralement au lieu de les « précéder » comme dans l'approche normative, qui fait intervenir un concept « transcendant » de signification.

Les autres affirmations saillantes ne sont finalement que des développements ou des conséquences de cette première décision. La décision de proposer un compte rendu empirique plutôt qu'une théorie normative motive en effet à son tour (A2) *le choix de s'en tenir aux « observables »*, et donc d'adopter une perspective en *troisième personne*, les observables devant être pour le moins *accessibles publiquement*, et *réitérables*. La question est alors (A3) de *déterminer ces observables*, et c'est tout naturellement vers Quine que l'auteur se tourne, notamment vers l'expérience de pensée de la *traduction radicale* (TR), qui reste l'expérience la plus aboutie de tentative de construction d'une théorie *empiriste* et *naturaliste* de la signification. Les observables pertinents sont donc déterminés à partir des contraintes de la TR, qui permet d'isoler comme seules données objectives assignables (p. 105) *les stimuli proximaux et le comportement verbal qui leur est causalement associé*. De là découle (A4) la détermination de *l'unité de compte de l'analyse* : la signification-stimulus des expressions ou phrases de la langue (indigène), que les contraintes de la TR doivent permettre de préserver (p. 139 et suivantes).

On pourrait penser qu'avec la signification-stimulus, dans la version amendée que propose M. Montminy (à juste titre, selon moi), qui la fait dépendre, avec Davidson, de stimuli *distaux* plutôt que *proximaux*, on soit finalement en possession d'un concept strict de signification. Mais ce n'est pas le cas. Car cela reste encore insuffisant pour déterminer ultimement la signification des occurrences verbales des locuteurs dans des circonstances déterminées, du fait qu'on n'est toujours pas en mesure de distinguer, dans la compréhension de chaque phrase, ce qui relève des stimuli distaux et ce qui relève de l'information collatérale et, donc, de faire la part, dans l'énoncé, de ce qui relève de sa composante *sémantique* et de ce qui relève des composantes *factuelle* et *doxastique*. L'indétermination de la traduction (IT) joue donc (A5) jusque dans les *phrases occasionnelles d'observation*, et apparaît *irréductible*. Cette irréductibilité n'est toutefois pas dommageable d'un point de vue « pragmatique » : elle n'affecte pas la communication et la compréhension. Sa seule incidence est en fait *épistémologique* : elle nous invite à *renoncer* à un concept strict de signification, et à une distinction stricte entre l'analytique et le synthétique.

3. Examen de quelques difficultés

3.1 L'unité de compte de l'analyse

Ma première remarque portera sur l'unité de compte de l'analyse, telle que la détermine l'approche empiriste. En reconstruisant l'argumentation précédente, on a mieux perçu les raisons des choix qui conduisent à isoler la signification-stimulus. Mais ces décisions imposent à l'analyse des limitations drastiques. À limiter le comportement verbal des locuteurs au fait d'accepter ou refuser un énoncé en certaines circonstances, on souscrit sans doute aux exigences de l'empirisme, mais on réduit l'objet accessible à bien peu de choses. M. Montminy reconnaît notamment qu'on laisse délibérément à l'écart la *force* et le *ton* (p. 65, n. 4), tout en admettant qu'ils sont finalement dispensables à l'analyse. Dès lors, il n'y aurait plus d'obstacle à prétendre reconstruire (via le manuel de traduction) la langue des locuteurs indigènes en examinant simplement les occurrences singulières d'expressions ou de phrases. Mais Putnam remarquait déjà³ qu'à procéder ainsi, la question se posait de savoir *ce que* le manuel de traduction permettait en fait de reconstruire : les règles de la langue indigène, ou les habitudes de parler des locuteurs, ou de certains d'entre eux? À cette première difficulté s'en ajoute une autre : à procéder ainsi, on perd de vue qu'une langue est par nature *colloquiale*, et que ce que les locuteurs ont à traiter lorsqu'ils *communiquent*, ce ne sont pas seulement des expressions ou des phrases, mais *des énonciations acquérant en contexte de communication valeur d'actes de langage*.

Dès lors, il est clair que ce que capture la « signification-stimulus » n'est qu'une partie de ce qui s'effectue comme sens dans l'instance d'énonciation, puisque se trouve seule retenue et prise en compte la composante *propositionnelle* des énonciations, elle-même réduite, dans le cas des phrases d'observation, à la seule composante *dénotationnelle*. Quand bien même on accorderait à une phrase une signification déterminée (fût-ce seulement sous la forme de la signification-stimulus), il resterait encore une source d'indétermination liée à l'*acte de langage* qu'elle permet d'accomplir en contexte. Il y va bien de la détermination de l'*unité de compte de l'analyse sémantique*, qui n'est jamais l'expression ou la phrase d'un locuteur isolé, mais bel et bien un *acte de langage en situation de communication*. Car, sauf à limiter (et encore) les phrases du langage indigène à la seule classe des assertions, *on ne peut considérer que les verdicts des locuteurs s'appuient sur la seule composante propositionnelle/dénotationnelle des énonciations*. Dans un grand nombre de cas, c'est au contraire sur la composante *illocutoire* de ces dernières que les verdicts sont basés. En effet, pour répondre par « oui » ou par « non » à la question : « En disant A, x a-t-il donné un ordre

3. Putnam, H., « The Analytic and the Synthetic », *Mind, Language and Reality*, vol. 2, Cambridge, Cambridge University Press, 1975, p. 33-69.

à y? », il faut que le locuteur interrogé reconnaisse s'il s'agissait ou non d'un ordre, et ainsi pour tous les actes de langage performables dans sa langue.

Certes, la mise à l'écart de cette composante illocutoire est justifiée par la perspective empiriste retenue : à vouloir s'en tenir aux *observables* accessibles dans une perspective en *troisième personne*, on est contraint de laisser à l'écart tout ce qui ne se laisse pas reconduire au comportement observable et aux stimuli proximaux ou distaux par une relation de dépendance causale. Sont alors écartés par décision méthodologique les connaissances d'arrière-plan, les intentions, les présuppositions, les univers de croyance, etc. Mais il devient du même coup impossible de déceler, dans le comportement linguistique et plus généralement communicationnel des locuteurs, si un ordre a été donné (et suivi), une promesse contractée (et tenue), etc.

On dira, à juste titre d'ailleurs, que les contraintes drastiques qu'impose la perspective empiriste sont susceptibles de s'assouplir. Chez Quine déjà, les conditions initiales de la TR, qui placent (idéalement) le traducteur dans la *même* situation que le locuteur natif qui apprend son langage, s'assouplissent progressivement. Si l'apprentissage s'inaugure avec des phrases occasionnelles d'observation, il n'en reste pas à ce seul et très élémentaire niveau. Le traducteur — comme l'enfant — apprend bientôt à *opérer des substitutions* dans les structures déjà connues et, de là, en vient progressivement à la maîtrise de sa langue par le biais de l'*interanimation des énoncés*, stade où n'est même plus requise la relation causale directe entre la situation et l'énoncé⁴. De même, M. Montminy montre bien que les contraintes de la TR n'opèrent véritablement que si on leur adjoint le Principe de Charité (p. 127 et suivantes) et la « Contrainte de la Communauté » (p. 137 et suivantes), ce qui réintroduit dans l'analyse à la fois les *croyances* et, dans une certaine mesure, la dimension *colloquiale* du langage⁵. Mais cela reste insuffisant, car la seule maîtrise du système de la langue, jointe au comportement observable de mon interlocuteur, *ne suffit pas* à me faire comprendre que par son énonciation, il est en train, par exemple, de me faire une promesse. Pour le comprendre, il me faut encore (i) déterminer que par son énonciation il *contracte l'engagement* d'accomplir l'acte futur décrit par l'énoncé (but illocutoire) ; (ii) *présupposer* qu'il le *sait*, en a la *capacité* (conditions préparatoires), et en a aussi l'*intention* (condition de sincérité) ; (iii) présupposer encore qu'il *sait* que la promesse ne sera *tenue* (conditions de satisfaction) que s'il accomplit lui-même l'acte futur décrit par l'énoncé (condition sur le contenu), autant d'éléments pour la détermination

4. Cf. Gochet, 1994, p. 334-335.

5. À vrai dire, les amendements qu'après Davidson, M. Montminy introduit dans la conception quiniennne mettent en évidence la dimension *sociale* (voir en particulier p. 114 et suivantes) du langage, plutôt que sa dimension *colloquiale*. Par là, on insiste sur le fait que la langue constitue, pour les membres de la communauté linguistique, un patrimoine commun, mais sans aller jusqu'à admettre que c'est seulement dans une relation interlocutive que les significations sont produites comme résultat d'une activité conjointe des interlocuteurs. Je pense, avec F. Jacques, qu'il convient d'aller jusque-là.

desquels les seuls traits linguistiques et comportementaux sont inopérants. Certes, le rôle ici dévolu aux états mentaux (désirs, intentions, croyances) ne laisse pas de soulever de sérieuses difficultés. Mais j'ai montré, dans un autre travail⁶, qu'on n'était pas tenu, pour rendre compte de la *compréhension* et de la *communication*, d'adopter un schème intentionnaliste et mentaliste de la signification linguistique, et qu'une pragmatique interactionnelle élargie permettait de surmonter les restrictions drastiques de la conception empiriste sans verser pour autant dans le mentalisme.

3.2 Renoncer à, ou redéfinir la signification?

Toutefois, on pourrait ici m'objecter, suivant Gochet⁷, que si on prend le *holisme* au sérieux, c'est la langue (ou la théorie) tout entière, ou le fragment de langue (ou de théorie), qui est l'*unité de signification*. La langue (théorie) articule des énoncés de différents niveaux (cf. la métaphore quiniennne du « centre » et de la « périphérie ») et de différents statuts (phrases occasionnelles, perdurables, d'observation, etc.), et ce sont les *relations* entre ces différents énoncés (ce que Quine nomme « l'interanimation ») qui *déterminent* la signification. Il est dès lors illusoire de parler de « la » signification d'un énoncé comme de quelque chose qu'on pourrait isoler, individuer, traduire, etc.

Je prendrai cette objection au sérieux, car ce sera pour moi l'occasion de préciser certaines autres réserves que j'entretiens à l'endroit d'un empirisme strict et, par contrecoup, de préciser d'autant ma propre conception. La question me semble être de savoir quelle conclusion on peut ou doit tirer des deux thèses du holisme et de l'IT, d'ailleurs complémentaires à certains égards. Car si le holisme nous dit qu'en toute rigueur, il est *incorrect* de parler de « la » signification d'un énoncé *isolé*, la thèse d'IT, de son côté, nous incite à renoncer à tenter de déterminer de façon ultime « la » signification d'un énoncé quelconque. Mais ce que nous dit finalement la thèse d'IT, dans sa version IS notamment, c'est que, même à un énoncé d'observation, je puis associer différentes significations sans pouvoir trancher : il y aura toujours différentes traductions de cet énoncé, *adéquates empiriquement* (i.e. en accord avec le comportement observable des locuteurs compétents et les stimuli distaux accessibles) mais cependant *incompatibles* entre elles. De là, on peut certes conclure que la signification n'est pas une entité associée à un énoncé. Mais de là à dire qu'il en irait de la signification comme il en allait autrefois des sorcières...

Peut-être qu'on ne peut déterminer ultimement « la » signification d'un énoncé tant qu'on le traite comme isolé. Peut-être que cette impossibilité tient à notre impuissance à déceler et à démêler la part du sémantique, du factuel et du doxastique dans la signification. Reste que *quelque chose comme la signi-*

6. Voir Grillo, Éric, *Intentionnalité et signifiante : une approche dialogique*, Bern, Peter Lang, 2000.

7. Gochet, 1994, p. 330.

fication se détermine et se spécifie entre nous, fût-ce rétrospectivement, dans l'espace de l'interlocution. Sans quoi il n'y aurait plus même de sens à parler de *compréhension* et de *communication*, car le simple constat d'habitudes verbales et comportementales communes et récurrentes au sein de la communauté linguistique ne permet pas d'affirmer que quelque chose a été *rendu commun*, ce qui est pourtant le cœur de ce qu'on nomme « communication »...

Si l'indétermination ne nuit pas à la compréhension, ce que j'accorde à M. Montminy, mais pour d'autres raisons que lui, sur lesquelles je reviendrai tout à l'heure, c'est que *la communication a justement pour tâche de la lever*, au moins partiellement et relativement. Si on prend au sérieux la communication linguistique, on doit admettre que c'est au sein d'une relation interlocutive que s'effectue la mise en communauté⁸ du sens, de la référence et de la force, qu'on nomme proprement *compréhension*. Le *processus de signifiante* n'a d'autre lieu où s'effectuer qu'une *relation interlocutive effective*, spécifiée dans une *stratégie discursive* particulière. Dès lors, on peut voir qu'une conception *dynamique* de la signification ne fait pas nécessairement appel à un concept « transcendant » de signification : « *la* » signification n'est que le *produit sédimenté et provisoirement fixé d'un processus de signifiante qui, ramené à ses conditions constitutives, articule les trois axes de la différence, de la référence et de l'interlocution*. « La » signification, c'est ce qui se construit et se spécifie entre nous, ici et maintenant, sur les trois axes de la signifiante, et au sein d'une relation interlocutive déterminée par une stratégie particulière. Sans doute chaque énoncé déploie-t-il une pluralité de significations possibles, mais entre elles, les interlocuteurs « choisissent » en fonction, d'une part, de leur *maîtrise commune de la langue*, et d'autre part, de leur *maîtrise conjointe des déterminants pragmatiques de la communication*. Sans doute peut-on concéder que la « signification-stimulus », chère aux empiristes, isole quelque chose comme un « noyau » de signification, mais d'autres paramètres encore interviennent dans la détermination de la *signification communiquée*, y compris pour les phrases d'observation. La tâche du traducteur radical, et les contraintes qu'il s'impose, lui permettent certainement d'isoler ce premier noyau, mais *seulement* ce premier noyau. Il en résulte que c'est dans la situation de la traduction radicale, paradoxalement, que l'*indétermination* est à la fois *maximale* et *irréductible*. Dans les circonstances usuelles de la communication, l'indétermination qui entache chaque énoncé au moment de son énonciation peut être rétrospectivement et interlocutivement levée. Si elle ne disparaît pas complètement, l'indétermination dans l'interlocution voit ses effets singulièrement atténués.

On aura repéré dans ce qui précède que je tombe d'accord avec M. Montminy pour reconnaître l'innocuité de l'IT à l'égard de la compréhension et de la communication. Mais il me semble que dans les limites d'un empirisme

8. Voir Jacques, F., « La mise en communauté de l'énonciation », *Langages*, 70, 1983, p. 47-71.

strict, on ne peut établir cette innocuité de manière tout à fait satisfaisante. Si l'IT « descend » jusqu'aux phrases occasionnelles d'observation, et M. Montminy l'a bien montré, on ne voit pas que les locuteurs indigènes soient mieux armés que le traducteur radical pour se prononcer sur « la » signification d'une telle phrase. C'est alors vers les habitudes communes de parler qu'il faut se tourner pour lever partiellement l'indétermination. Mais comme l'a bien souligné Putnam, on risque alors de *confondre* habitudes de parler et règles du langage, au risque alors d'*accentuer*, au sein de la communauté linguistique elle-même, les différences entre idiolectes et sociolectes, jusqu'à faire de la communication un malentendu généralisé. Ainsi, un locuteur français peut-il être déconcerté la première fois qu'il entend la tournure, pourtant fort usitée dans le Sud-Ouest : « Je te promets qu'elle est venue », qui ne correspond pas à l'usage « canonique » de « promettre » en français, et qui, de plus, illocutoirement, n'est pas un engageant, mais un assertif, équivalent à « je t'assure que... ». Sans doute ne peut-on associer *in abstracto* de signification déterminée à cet énoncé, pas plus qu'à un autre d'ailleurs. Sans doute différentes hypothèses analytiques motiveraient-elles des traductions empiriquement acceptables, quoique incompatibles entre elles : c'est le nerf de la thèse d'IT. Mais alors, de deux choses l'une : soit on conclut qu'il n'y a rien de tel que la signification, mais alors, l'*indétermination* pourrait à terme signifier l'*incommensurabilité*, et on ne voit plus comment on pourrait encore parler de compréhension et de communication en donnant à ces termes la signification intuitive que notre expérience quotidienne nous permet de leur associer. Soit on conclut qu'indétermination *n'est pas* incommensurabilité, et que, si on ne peut associer *in abstracto* à une expression une signification déterminée, c'est que *la signification n'est déterminable qu'interlocutivement*.

Tant que le traducteur radical n'est pas en mesure d'entrer avec les indigènes dans une relation interlocutive effective, différents manuels restent possibles, empiriquement acceptables, mais mutuellement incompatibles. Mais les choses *changent* quand cette possibilité est offerte : non pas que l'on puisse désormais départager les manuels en correct/incorrect du point de vue des comportements linguistiques observables des locuteurs, mais on peut désormais amender chaque manuel *en éliminant les traductions que l'échange interlocutif révèle manifestement incorrectes*. Je reviens à l'exemple précédent : mon embarras s'estompe, dès lors qu'à ma question : « Comment peux-tu promettre quelque chose qui a déjà eu lieu et que tu n'as pas toi-même accompli? », je m'entends répondre : « Mais ce n'était pas une promesse. » Le point étant qu'alors, *ce ne sont pas les observables qui changent*, mais bel et bien les *hypothèses analytiques* dont certaines s'avèrent désormais fausses ou superflues.

« La » signification se détermine en se construisant entre nous dans l'espace interlocutif, jusqu'à un certain point. Vouloir atteindre un degré « ultime » de détermination est à la fois impossible théoriquement et inutile communicationnellement. Impossible théoriquement, car le contraire vou-

draît dire que la signification est déterminable intégralement, *indépendamment* du contexte interlocutif où elle s'effectue, ce qui est inexact, y compris pour des énoncés « analytiques » d'ailleurs, puisqu'un énoncé tel que : (x) $(x = x)$ peut, selon son contexte d'emploi, être interprété comme loi sémantique ou comme loi physique. Inutile communicationnellement, car tout ce dont on a besoin dans la communication, c'est de spécifier la signification *en contexte* de l'énoncé, dans le *moment* de son énonciation.

Il me semble que les amendements que M. Montminy apporte à l'expérience quinquienne de la traduction radicale, que ce soit le Principe de Charité ou la Contrainte de Communauté, manifestent qu'il a conscience de ces difficultés et qu'il reconnaît à la signification une dimension *relationnelle*. Mais la relation, ramenée simplement à des *habitudes communes* de comportement linguistique⁹, reste *extérieure* au procès du sens, agissant comme simple principe régulateur pour la communication et la compréhension. Je considère pour ma part qu'elle joue un rôle *constitutif* dans la *détermination* de la *signification communiquée*.

Si on accorde ce qui précède, on voit qu'il est possible de défendre une théorie *dynamique* de la signification qui tout à la fois : (i) reste empiriquement acceptable, en tant qu'elle *s'appuie* sur les pratiques linguistiques effectives, au lieu de prétendre les réguler ; (ii) ne fait pas intervenir un concept « transcendant » de signification, puisqu'au contraire on admet que la signification communiquée ne se détermine *qu'au sein* d'une relation interlocutive effective ayant cours entre nous ; (iii) enfin, atténue les effets de l'IT, tant au plan intra linguistique qu'au plan inter linguistique, les protagonistes de la communication effectuant à mesure les ajustements (sémantiques et doxastiques) requis pour la poursuite de la communication, sous la juridiction de l'objectif commun.

4. Pour un empirisme non réductionniste

Voilà une conclusion qui semble moins « pessimiste » que celle de M. Montminy. La stratégie empiriste qu'il adopte le mène en effet à ce résultat qu'*un concept strict de signification n'est pas à notre portée*, faute de pouvoir jamais être suffisamment déterminé et correctement spécifié. Si les notions d'analyticité, de synonymie, de traduction restent obscures, c'est parce que nous ne disposons pas d'un concept adéquat de signification. Mais l'adéquation ici visée est celle que définit une stratégie empiriste soucieuse, d'une part, d'en rester aux observables et, d'autre part, de fournir, pour chaque distinction ou chaque concept introduit, une caractérisation en termes de conditions nécessaires et suffisantes, réputée seule acceptable. C'est parfaitement clair dans le cas de la synonymie :

9. Ce qui semble attesté par la caractérisation que donne M. Montminy de la « Contrainte de Communauté » d'une part (p. 99 et suivantes) et de ce qu'il nomme la « stimulus-synonymie communautaire » de l'autre (p. 148).

Insistons encore une fois sur le fait que nous sommes à la recherche d'un critère de synonymie au sens fort du terme : on doit spécifier les liens nécessaires et suffisants qui doivent être réalisés entre deux formes linguistiques pour qu'on puisse correctement les décrire comme synonymes (p. 29).

Et il en va de même pour l'analyticité et la traduction : dans le premier cas, on veut isoler ce qui, dans le jugement, relève de la signification, de ce qui relève d'éléments factuels ou doxastiques ; dans le second, on veut déterminer ce qui, dans la traduction, est préservé au titre de « signification », dont on a vu d'ailleurs que, pour l'empiriste, ce ne pouvait guère être autre chose que la signification-stimulus. Ces contraintes drastiques n'étant pas satisfaites, *c'est le concept même de signification qui devient suspect* et dont on doit se départir (p. 207). Si la stratégie empiriste échoue, c'est donc, finalement, faute d'objet.

Mais il me semble, une fois encore, qu'une autre réponse est possible, moins pessimiste, et plus « libérale », qui consisterait à dire que si la stratégie empiriste échoue, dans la version au moins qui a été défendue par l'auteur, c'est peut-être qu'elle est inappropriée, parce que réductrice, et qu'elle n'honore pas toute la complexité de l'objet qu'elle envisage : la signification.

J'ai déjà eu l'occasion plus haut de montrer que la stratégie empiriste, parce qu'elle néglige volontairement la composante illocutoire de la signification, est réductrice en un premier sens. Les remarques qui précèdent me laissent penser qu'elle l'est encore en un second sens, car elle néglige ce que Benveniste nommait « l'articulation sémantique¹⁰ ». Pour le dire en un mot, il se pourrait que les critères d'identification et d'individuation des « significations » ne soient pas déterminables en termes de conditions nécessaires et suffisantes, mais plutôt dans des termes voisins des « airs de famille » par quoi Wittgenstein caractérisait ses « jeux de langage », et que ce fait soit directement lié, justement, à « l'articulation sémantique ».

Ce que mettait en exergue Benveniste avec ce concept, c'est que la langue fonctionne selon *deux plans d'organisation distincts, mais articulés* dans l'acte de signifier. Laissons ici la parole à Benveniste :

Or, les mots, instruments de l'expression sémantique, sont, matériellement, les « signes » du répertoire sémiotique. Mais ces « signes », en eux-mêmes conceptuels, génériques, non circonstanciels, doivent être utilisés comme « mots » pour des notions toujours particularisées, spécifiques, circonstancielle, dans les acceptions contingentes du discours. Cela explique que les signes les moins délimités à l'intérieur du répertoire sémiotique de la langue, « être », « faire », « chose », « cela », aient, comme mots, la plus haute fréquence d'emploi¹¹.

En passant du signe au mot, on passe de la langue comme système (plan sémiotique) à la langue comme instrument de communication (plan sémantique¹²),

10. Voir Benveniste, É., « La forme et le sens dans le langage », dans *Problèmes de linguistique générale*, t. 2, Paris, Gallimard, 1983, p. 214-238, notamment p. 228-229.

11. *Ibid.*, p. 228.

12. On dirait aujourd'hui, plus justement sans doute, plan sémantico-pragmatique.

via l'événement de l'énonciation. Dans ce passage, le signe perd sa valeur générique pour acquérir, comme mot, une valeur d'emploi, déterminable comme *sens communiqué*. De là l'impossibilité d'une traduction *radicale* qui voudrait que dans la traduction « la » signification soit préservée : chaque langue effectuant en son sein, sur un mode spécifique, l'articulation de ces deux plans d'organisation, les recouvrements, s'ils sont possibles, ne sauraient jamais être que partiels. Ce que Benveniste d'ailleurs, avait parfaitement aperçu :

On peut transposer le sémantisme d'une langue dans celui d'une autre, « *salva veritate* » ; c'est la possibilité de la traduction ; mais on ne peut pas transposer le sémiotisme d'une langue dans celui d'une autre, c'est l'impossibilité de la traduction. On touche ici la différence du sémiotique au sémantique¹³.

Voilà pourquoi un manuel de traduction ne ressemblera jamais à une simple liste de paires de mots, comme un dictionnaire bilingue par exemple, ni à une simple liste de paires de phrases, comme on en trouve dans ces guides touristiques qui prétendent, en quelques phrases-clés, nous mettre en mesure de nous sortir des situations les plus usuelles que le touriste-type aura nécessairement à affronter, et dont il n'est un secret pour personne qu'elles sont le plus souvent source d'inoubliables quiproquos...

Maîtriser une langue, c'est donc, avant tout, maîtriser cette articulation sémantique qu'elle réalise sur un mode spécifique, et dont on peut douter qu'elle soit accessible à celui qui voudrait simplement se borner, dans son apprentissage, au comportement verbal observable des locuteurs et aux liens causaux qu'il entretient avec les stimuli distaux pertinents.

5. Remarques conclusives

Si on accorde ce qui précède, on comprendra mieux ce que j'entendais en commençant, lorsque je qualifiais ma position de « modérée » et de « libérale », pour la distinguer d'un « empirisme strict ». Certes, il convient de *lier* une théorie sémantique, quelle qu'elle soit, à des données empiriquement accessibles qui l'étayent, car c'est là la meilleure garantie contre la tentation des vaines conjectures. Mais il convient de se garder aussi des critères qu'on adopte dans l'analyse : à les vouloir trop exigeants, on risque, sous couvert de rigueur, de tomber dans un *réductionnisme* dommageable. C'est la raison pour laquelle je me refuse pour ma part à conclure, de l'échec de la stratégie empiriste de fondation du concept, au caractère illusoire ou chimérique de la signification. Je pense enfin que, sur des questions aussi fondamentales, on ne peut espérer d'avancées décisives que si logiciens, philosophes et linguistes travaillent main dans la main. Mais cela est une autre histoire.

13. *Ibid.*

Réponse à Éric Grillo

Martin Montminy

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004958ar>
DOI : <https://doi.org/10.7202/004958ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Montminy, M. (2000). Réponse à Éric Grillo. *Philosophiques*, 27 (1), 203–206.
<https://doi.org/10.7202/004958ar>

Réponse à Éric Grillo

MARTIN MONTMINY

montminy@uottawa.ca

Département de philosophie

Université d'Ottawa

Dans mon livre *Les fondements empiriques de la signification*, j'explore la question de savoir quelles sont les données empiriques pertinentes pour construire une théorie de la signification et jusqu'à quel point ces données permettent de déterminer la signification. Éric Grillo résume très bien les grandes étapes de ma démarche. Il émet toutefois des réserves concernant certaines des thèses que je défends. Je vais brièvement présenter ces thèses et montrer qu'elles ne sont pas affectées par les critiques de Grillo.

1. Empirisme strict et empirisme modéré

Grillo affirme déceler deux versions de l'empirisme dans mon livre, qu'il appelle respectivement « empirisme strict » (EM) et « empirisme modéré » (ES). Ces deux thèses affirment :

(EM) Une théorie sémantique légitime devrait lier tout énoncé à propos de la signification à des données empiriques en principe accessibles.

(ES) Les questions fondamentales en épistémologie et en métaphysique ne peuvent s'appuyer sur une notion bien définie de signification.

Grillo tient EM pour peu contestable, mais estime que ES est problématique. La difficulté n'est pas que EM et ES sont incompatibles. (Les expressions qu'il choisit pour désigner les deux thèses sont donc quelque peu trompeuses, mais passons.) En fait, Grillo soutient que EM n'entraîne ES que si l'on admet un point de vue extrêmement limité sur la nature des données empiriques acceptables pour fonder la théorie sémantique. J'aborderai cette question à la prochaine section. J'aimerais pour l'instant expliquer pourquoi j'affirme que ES ne remet pas en question nos capacités à communiquer et à nous comprendre mutuellement.

Considérons la situation dans laquelle un tenant de la théorie relativiste (appelons-le Albert) doit interpréter quelqu'un qui adhère à la mécanique classique (appelons-le Isaac)¹. À quoi réfère le mot « masse » tel qu'utilisé

1. Cet exemple s'inspire de la discussion de Field, 1973. Voir Montminy (1998, p. 177-182) pour plus de détails et pour d'autres exemples d'indétermination de la vérité.

par Isaac? Il semble qu'il y ait (au moins) deux réponses possibles : ce mot peut référer à la masse relativiste ou à la masse au repos. Si tel est le cas, il y a indétermination de la traduction de la langue d'Isaac dans celle d'Albert ; plus précisément, il y a ce que j'appelle « indétermination de la vérité », puisque selon l'interprétation choisie, l'énoncé « La masse est invariante d'un système inertiel à un autre », accepté par Isaac, sera interprété par Albert comme étant vrai ou faux. L'indétermination empêche-t-elle Albert de comprendre Isaac? Certainement pas. Si Albert connaît les principes centraux de la théorie d'Isaac, il peut connaître les dispositions verbales de ce dernier, et cela est en principe suffisant pour le comprendre : le choix particulier d'un manuel de traduction n'ajoute rien à sa compréhension d'Isaac. Il n'est donc pas nécessaire de poser l'existence d'une signification qui serait communiquée d'Isaac à Albert (ou de « quelque chose qui a été rendu commun », selon les mots de Grillo) lorsque celui-ci comprend les énoncés de celui-là. La morale est qu'une notion déterminée de signification n'est pas requise pour expliquer la communication et la compréhension mutuelle, et les affirmations alarmistes de certains commentateurs de Quine concernant les conséquences absurdes ou désastreuses de l'indétermination de la traduction ne sont nullement justifiées.

Quelles sont donc les conséquences de l'indétermination? Les mêmes que celles qui découlent du rejet de la distinction analytique-synthétique. Ces conséquences peuvent être succinctement exprimées par ES, qui affirme qu'il est vain de tenter de justifier une position sur le statut épistémologique de certains énoncés ou sur la nature de certaines entités ou propriétés du monde en se fondant sur une notion déterminée de signification. Considérons encore une fois l'énoncé « La masse est invariante d'un système inertiel à un autre » accepté par Isaac. S'agit-il d'un énoncé analytique, c'est-à-dire vrai uniquement en vertu de la signification? Si les remarques qui précèdent sont justes, cette question n'a pas de réponse déterminée. Ainsi, tenter d'établir le statut épistémologique de cet énoncé (est-il un énoncé *a priori*, c'est-à-dire un énoncé pouvant être justifié simplement en s'appuyant sur la signification et indépendamment des faits?) est analogue à tenter de découvrir des sorcières : les soi-disant preuves ou découvertes sont sans fondement.

2. Signification cognitive et communication

Quine s'appuie sur les dispositions des locuteurs à accepter ou refuser des phrases dans diverses circonstances observables (ci-après, les dispositions verdictives) pour rendre compte de la signification de ces phrases ou, pour être plus précis, de ce que l'on pourrait appeler leur « signification cognitive ». (Je vais suivre ici l'usage en vigueur dans mon livre et employer le mot « signification » pour désigner la signification cognitive.) Comme le souligne Grillo, les données relatives aux dispositions verdictives ne permettent d'expliquer que l'un des multiples aspects de la communication entre

locuteurs. Si l'on désire une explication complète de la communication, il faut aussi tenir compte (entre autres choses) des données relatives à l'aspect illocutoire (ou force) des énoncés. Un compte rendu de la signification d'un énoncé donné ne nous dit pas s'il s'agit d'une promesse, d'une question, d'un ordre, etc. Il va donc de soi que la signification ne peut suffire à expliquer tous les aspects de la communication, et je n'affirme certainement pas le contraire dans mon livre. Tout ce que je propose, c'est de me concentrer sur la question plus restreinte de la composante cognitive de la signification.

Je ne vois toutefois pas en quoi ces considérations remettent en question ES et la thèse de l'indétermination de la traduction. Grillo suggère qu'une théorie de la signification peut être enrichie de données relatives à la composante illocutoire, et que celles-ci permettent en fait de lever (en partie) les indéterminations admises par les interprétations construites sur la base des seules dispositions verdictives. Je n'aborde pas vraiment cette question dans mon livre, mais il me semble douteux que cela soit le cas. Voyons pourquoi.

Tout d'abord, il n'est certes pas évident qu'une théorie de la force pourrait être formulée indépendamment d'une théorie de la signification. Mettons toutefois ce problème de côté, et supposons qu'il soit possible de rendre compte de la composante illocutoire sans présupposer une connaissance de la signification. Supposons donc que les diverses attitudes associées à la promesse, l'ordre, l'interrogation, etc. puissent être identifiées indépendamment de la signification des énoncés impliqués, autrement dit, que ces attitudes soient en quelque sorte du même type que l'attitude (davidsonienne) de tenir un énoncé pour vrai. Il serait par exemple possible de déterminer qu'en produisant l'énoncé *S*, le locuteur pose une question, sans avoir à connaître la signification de *S*. La question est de savoir si l'identification des différents actes de langage accomplis par les locuteurs autorise des distinctions de signification plus fines que celles que permettent d'établir leurs dispositions verdictives. Je n'ai rien vu dans le commentaire de Grillo qui motive une réponse affirmative à cette question.

Mes réserves à l'endroit de la suggestion de Grillo pourraient être exprimées de la façon suivante : en quoi le fait de savoir que le locuteur nous pose une question lorsqu'il produit *S* nous éclaire-t-il sur la signification de *S*? Grillo pourrait peut-être proposer d'exploiter ce genre de données de la même façon que Davidson exploite les données relatives à l'attitude de tenir des énoncés pour vrais. Ainsi, pourrait-il souligner, savoir que le locuteur pose une question en énonçant *S* nous permet d'inférer un certain nombre de choses sur son état d'esprit, par exemple qu'il ne sait pas si *S* est vrai. La difficulté est que rien n'indique que ce genre d'information ne peut aussi être obtenu sur la base des dispositions verdictives du locuteur : on peut en effet inférer que le locuteur ne sait pas si *S* est vrai à partir de sa disposition à donner un verdict d'abstention si on lui propose *S*. En général, un locuteur qui accomplit un certain acte illocutoire possède aussi des dispositions verdictives qui peuvent nous informer sur les présuppositions de cet acte (et sur

bien d'autres choses). Grillo semble sous-estimer la richesse des données fondées sur les dispositions verdictives.

Ces considérations peuvent être illustrées à l'aide de l'exemple du locuteur français du Sud-Ouest, dont l'usage de l'expression « Je te promets » ne correspond pas au nôtre. Il importe de noter qu'on peut aisément voir dans ce cas que le manuel de traduction homophonique doit être rejeté simplement en considérant les dispositions verdictives de ce locuteur concernant les phrases impliquant cette expression. J'ajoute tout de suite qu'il ne faudrait pas penser que l'exclusion du manuel de traduction homophonique élimine toute forme d'indétermination. Comme j'ai essayé de le montrer dans mon livre, l'indétermination de la traduction est liée à l'absence de critère permettant de séparer complètement la signification de la croyance (factuelle) ; Grillo n'a pas montré en quoi le recours à la composante illocutoire permettait d'effectuer cette séparation. L'aisance avec laquelle nous pouvons exclure certaines traductions inadéquates dans des cas comme celui du locuteur du Sud-Ouest est faussement rassurante.

Références

- Field, H., 1973. « Theory Change and the Indeterminacy of Reference », *Journal of Philosophy*, 70, p. 462-481.
- Montminy, M., 1998. *Les fondements empiriques de la signification*, Montréal/Paris, Bellarmin/Vrin.

Philosophiques

philosophiques

Alain Badiou, *Court traité d'ontologie transitoire*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 199 p., *Petit manuel d'inesthétique*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 214 p. et *Abrégé de métapolitique*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 166 p.

Alain Beaulieu

Le matérialisme contemporain

Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004897ar>

DOI : <https://doi.org/10.7202/004897ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Beaulieu, A. (2000). Compte rendu de [Alain Badiou, *Court traité d'ontologie transitoire*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 199 p., *Petit manuel d'inesthétique*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 214 p. et *Abrégé de métapolitique*, Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 166 p.] *Philosophiques*, 27 (1), 207–211. <https://doi.org/10.7202/004897ar>

Tous droits réservés © Société de philosophie du Québec, 2000

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>

Comptes rendus

Alain Badiou, *Court traité d'ontologie transitoire*. Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 199 p. Alain Badiou, *Petit manuel d'inesthétique*. Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 214 p. Alain Badiou, *Abrégé de métapolitique*. Paris, Seuil, (coll. « L'ordre philosophique »), 1998, 166 p.

Dans les trois livres qu'il nous présente, Badiou récapitule et consolide les acquis de *L'Être et l'événement* (Seuil, 1988), où était mise en place l'architecture de son système (cf. aussi *Manifeste pour la philosophie*, Seuil, 1992). L'auteur poursuit son travail de refondation de la philosophie à travers de savantes méditations sur la science mathématique, les arts de la scène et de l'écriture, de même que la politique. Ces trois champs d'investigation réfèrent aux principaux axes de recherche de Badiou où s'effectuent les « procédures de vérité ». L'amour comme quatrième procédure, précédemment placé sur le même plan que les trois autres, a momentanément été laissée en suspens. Avant d'aborder ces plus récents textes, revenons sur les caractéristiques générales du système de Badiou en nous demandant de quelle façon le champ philosophique est circonscrit par les quatre sphères de production de la vérité. L'auteur appuie son système sur la philosophie de Platon. L'entreprise platonicienne, dont chacun des moments se ramène, selon Badiou, à l'étude de l'une ou l'autre des quatre procédures de la vérité, aurait noué pour la première fois l'ontologie aux mathématiques. Savant esprit mathématique, Badiou associe sa pensée au platonisme d'où il exclut l'hypothèse de l'idéal unitaire du Bien à la faveur de l'apologie du multiple. La réhabilitation de Platon et du concept de vérité revêt, bien sûr, un ton provocateur en regard des tendances antifondationnelles de notre époque. Il ne s'agit pourtant pas de réaffirmer le plein pouvoir de la philosophie sur les autres domaines du savoir. Au contraire, la philosophie ne sera refondée qu'une fois déchargée de ses pouvoirs traditionnels qu'elle abandonne aux mains des mathématiques, de la poésie, de la politique et de l'amour, à qui seuls revient le privilège de produire des vérités. Badiou invite donc le philosophe à déconsidérer l'emprise de sa science sur la vérité et à expérimenter des vérités produites indépendamment de son intellection en se plaçant sous les conditions des quatre procédures déjà pleinement pensantes en elles-mêmes. C'est ainsi que Badiou se met à l'écoute de la pensée d'une impressionnante diversité d'auteurs, pour la plupart non consacrés comme philosophes, tout en ne manquant pas de dénoncer les figures de la philosophie qui prétendent détenir tous les droits sur la pensée.

1. *Court traité d'ontologie transitoire*

Le premier chapitre analyse une triple identification de Dieu : le Dieu de la religion, le Dieu de la métaphysique et le Dieu des poètes. Alors que le premier est mort et que la déconstruction du second est déjà bien entamée, le troisième se fait plus persistant et sert encore de guide non seulement aux poètes de cet âge, de Hölderlin à Celan, mais aussi à certains philosophes, tels que Heidegger, dont le rapport à la langue poétique cache une obscure quête de transcendance. Désirant en finir avec toutes les pro-

messes d'un retour des Dieux, Badiou place sa pensée sous le signe d'un athéisme radical. Comment alors fonder une ontologie qui ne soit plus dans la transcendance? Ou, ce qui revient pour lui au même, une philosophie qui n'associe plus l'être à l'un? La solution proposée par l'auteur consiste à faire un saut dans les mathématiques et à mettre la philosophie sous condition de la théorie des ensembles de Cantor, seule en mesure de penser le multiple pur ou le multiple sans un et dont l'un des axiomes principaux consiste à dire qu'il n'y a pas d'ensemble de tous les ensembles. Cette théorie, dont la formulation hypothétique de l'impératif ontologique exige que l'être soit pensé comme pur multiple, devient l'une des grandes avancées des mathématiques à la hauteur de laquelle la philosophie n'a pas encore su se placer. Le postulat de l'adéquation de l'ontologie à la mathématique, aujourd'hui tombé dans l'oubli, promeut ainsi la philosophie au rang d'un pur raisonnement axiomatique et évacue tout rapport de la philosophie avec la poésie de la transcendance, l'opinion, le jugement ou l'interprétation. L'être est toujours fondé absolument comme le sont les mathématiques dont le développement correspond à celui de l'ontologie.

La découverte du statut ontologique des mathématiques permet à l'auteur de donner une structure à la philosophie de notre époque qui, force est de l'admettre, a le plus souvent tendance à discréditer la notion de fondement en associant la philosophie à une sorte d'archi-esthétique ou, plus radicalement, à une forme d'irrationalisme. Badiou voit bien les limites de notre contemporanéité dans son geste presque généralisé de raccord entre la philosophie et la poésie. Notre principale tâche devient celle de défaire la plus solide suture encore persistante à notre époque entre la philosophie et le dire poétique. Le démontage de cette union coriace au profit de la rationalité mathématique devrait donner la chance à la philosophie de faire un bond en avant et permettre à l'ontologie de renouer avec son essence depuis toujours mathématique.

On voit défiler dans le *Court traité d'ontologie transitoire* une imposante liste des notions à combattre qui sont autant d'entraves au projet de refondation de la philosophie dans la science mathématique. On peut prendre la mesure du caractère révolutionnaire de l'entreprise de Badiou à l'inventaire qu'il fait des contresens produits et entretenus par notre philosophie. Ces errements propres à notre époque sont si nombreux qu'ils en viennent presque à se confondre avec l'entière des avancées marquantes de la pensée de notre siècle. Le succès espéré de leur mise en échec reconfigurerait donc totalement les données de la philosophie actuelle. Ce qui en dit long sur les ambitions de Badiou qui frôleront pour certains la démesure. Sans que l'on puisse ici entrer dans une explication détaillée de chacun des points particuliers de mécontentement de l'auteur en regard des tendances dominantes de la philosophie contemporaine, qu'il soit permis d'en donner quelques repères où nous soulignons à chaque fois les adversaires d'une pensée du multiple et du pur : la *finitude* et l'*herméneutique* restent aux prises avec le motif de l'un transcendantal dont seule peut nous sortir une pensée du multiple pur ou sans-un modelée sur la théorie des ensembles de Cantor (p. 21) ; l'éloge du *perspectivisme* et la soif du *consensus* demeurent prisonniers des jugements d'existence foncièrement impossibles à partager, tout en méconnaissant le caractère non impératif des mathématiques (p. 54) ; le *tournant langagier* dans ses versions heideggerienne et wittgensteinienne ignore le primat du mot sur la chose dans le langage mathématique (chap. 8) ; la *philosophie analytique* a porté au comble la mathématisation de la logique en oubliant le caractère ontologique des mathématiques (p. 125, 136) ; l'*idéalisation du langage poétique* ou la préférence de la catégorie du beau sur celle du vrai demeure sourde au fait que

les mathématiques puissent être autre chose que du calcul (p. 122 et ailleurs). Bref, l'essence de l'argument de Badiou contre la tentation non fondationnelle tient dans cet énoncé simple selon lequel l'ontologie est la mathématique elle-même. Ce qui était pleinement su par Platon, rappelle Badiou, puis occulté par Aristote, qui ne reconnaît toujours qu'un statut fictionnel et jamais ontologique ou pensant aux *matematika* (p. 185-186). Notre siècle, qui déconsidère le caractère de pensée des mathématiques, est à cet égard pleinement aristotélicien (p. 111).

2. *Petit manuel d'inesthétique*

Dans son *Petit manuel d'inesthétique*, l'auteur souhaite penser non pas *sur* mais *avec* les formes d'expressions artistiques (la danse, le théâtre et le cinéma) et les artistes, poètes ou écrivains pour la plupart (Pessoa, Celan, Mallarmé). On reconnaît dans ces choix les affinités de Badiou avec les arts de la scène et les arts de l'écriture, lui qui est également dramaturge et romancier. Il ne s'en tient heureusement pas à une simple confrontation des styles et s'adonne plutôt à un impressionnant jeu de transfiguration des écrivains en philosophes. À cela, il nous avait déjà habitués. Sans doute la plus profonde originalité du texte repose-t-elle dans la définition du concept d'inesthétique, terme générique sous lequel l'auteur place ici ses réflexions.

Qu'entendre par inesthétique? C'est ce à quoi répond le premier chapitre du livre où l'auteur y récapitule l'histoire de l'esthétique occidentale à l'aide de trois notions qui sont autant de types de nouage entre l'art et la philosophie. Il y a d'abord le schème didactique d'orientation platonicienne selon lequel l'art ne contient aucune vérité, celle-ci lui étant toujours extérieure. Vient ensuite le schème romantique en vertu duquel, au contraire du précédent, l'art est le producteur absolu de la vérité. Et en troisième lieu, le schème classique où l'art et la science sont en paix et d'où sont exclus tous les rapports de domination de l'un sur l'autre. On retrouve cette paisible cohabitation dans l'œuvre d'Aristote, mais aussi chez les rationalistes modernes (Descartes, Leibniz et Spinoza), qui se sont fort préoccupés d'assigner un statut distinct à l'art. L'auteur remarque que notre siècle a été incapable de produire un schème vraiment nouveau par rapport aux précédents, les trois plus originales interprétations esthétiques ayant été celles du marxisme, de la psychanalyse et de l'herméneutique allemande qui sont respectivement ramenées à des variantes tardives des schèmes didactique, classique et romantique. Quant aux avant-gardes du début du siècle, elles n'auraient fait que chercher un schème médiateur de type didactico-romantique. L'inesthétique propose donc un quatrième schème ou un nouveau type de rapport entre l'art et la philosophie. De même que l'ontologie a pu se définir par les mathématiques en leur reléguant son titre de pensée, l'inesthétique ne s'accorde plus, comme c'était le cas dans l'esthétique traditionnelle, un privilège de pensée sur l'art et considère plutôt le statut pensant de l'art lui-même. La philosophie n'a plus qu'à montrer la vérité produite par les œuvres d'art.

L'auteur considère les expressions artistiques comme des producteurs singuliers de vérité. La philosophie accuse toujours un retard par rapport aux vérités de l'art et il est de son devoir de les identifier pour se mettre à leur hauteur, sous leurs conditions. Pour ne donner que quelques exemples qui surgissent au cours des analyses et qui mettent en valeur le statut pensant de l'art, mentionnons l'œuvre de Pessoa qui arrive à penser la coexistence des univers sensible et intelligible sans retomber dans le despotisme de l'un en plus de penser des singularités multiples sans renouer avec

l'empirisme (p. 73) ; le travail de Beckett, à travers lequel s'élabore une profonde pensée du dire comme ratage, allant même jusqu'à espérer un ratage absolu en vue d'en finir une fois pour toutes avec le dire (p. 155) et dont l'absurde espoir se double aussi de l'expérience du courage qui naît de ce que les mots ont parfois tendance à sonner vrai (p. 163) ; et, surtout, située à l'autre extrémité de la langue et opposée à la théorie des ensembles de Cantor, l'œuvre de Mallarmé où l'on trouve une formidable pensée de l'événement pour laquelle le fameux « Coup de dés » demeure emblématique (chap. 5 et 10). Badiou nous annonce la parution prochaine de *L'Être et l'événement*, seconde partie, où il approfondira la nature du lien entre l'art et l'ontologie en tentant de donner une consistance nouvelle et non transcendante au point de croisement des univers mathématique et poétique.

3. *Abrégé de politique*

La métapolitique est un appel lancé au philosophe, l'invitant à placer son travail sous les conditions de la pensée politique. Qu'est-ce que cela implique? Rien de moins que la fin de la philosophie politique. Sans être inscrit dans un processus de militantisme, le philosophe qui se croit spécialiste de la politique en s'accordant sur elle les pleins pouvoirs d'analyse et en déterminant de l'extérieur les lois d'une politique idéale ne fait qu'entretenir une vague confusion. Seuls ceux qui ont une expérience politique ont quelque chose de déterminant à dire sur la politique. C'est en tant qu'ex-maoïste engagé et militant de longue date que Badiou prend position contre les régimes démocratiques ou capitalo-parlementaristes, qui ne sont pour lui qu'un festival d'opinions amenées sur la place publique, invariablement soumises à la voie majoritaire et incapables de générer une véritable remise en cause décisionnelle favorable à une réelle émancipation. Le parlementarisme, parce qu'il ne décide de rien, est une idéologie à craindre. Il fut abstraitement promu par la philosophie politique de Arendt, critiquée de façon acerbe par l'auteur (chap. 1).

Mettre la philosophie sous les conditions de la politique ne signifie pas qu'il faille la placer sous les conditions de n'importe quelle politique. Ce sont les politiques d'émancipation qui doivent servir de guide. Plus précisément, c'est sous les conditions du cycle révolutionnaire et prolétarien dont les principaux jalons sont les pensées de Marx, Lénine et Mao Tsé-toung, constitutifs du premier cycle des politiques modernes d'émancipation (p. 77) que la philosophie doit se placer. De là, l'auteur pose de réelles questions qui visent bien sûr à mieux comprendre la politique historique, mais aussi à ébranler le régime plus général des écoles philosophiques. Au nombre de ces interrogations, Badiou se demande pourquoi les guerres de libération débouchent si souvent sur de solides et impénétrables constructions étatiques (p. 79). En vue de se soustraire à cette logique de révolution/fortification, l'auteur s'interroge sur la nécessité pour un militant d'adhérer à un parti. Désirant trouver une alternative à la fois au communisme, pour lequel il n'entretient qu'une nostalgie partielle, et au régime capitalo-parlementariste, qui ne fait toujours qu'imposer une opinion majoritaire, l'auteur spéculé sur ce qu'il appelle une politique sans parti (p. 138).

Au sein d'une politique sans parti, on trouve des militants sans chef. Dans un style très deleuzien, l'auteur convie le philosophe à se joindre à la multiplicité affirmative pour faire naître un homme polyvalent et soustrait aux dictatures communistes ou parlementaristes de l'État. La figure du militant sans parti rappelle étrangement, sans qu'il en soit fait mention, le libre penseur de Spinoza (livre XX du *Traité théologico-*

politique) à cette différence près, fort appréciable cependant, que la liberté de penser du militant sans parti n'est jamais accordée au préalable au souverain. Le militant sans parti, qui devient le plus apte à prendre des décisions politiques, est donc celui qui ne prête allégeance à aucun souverain et à aucun État. Il n'attend plus rien des grands mouvements révolutionnaires et se connecte sur une multiplicité d'idéologies sans jamais en considérer une seule comme parfaitement achevée. L'auteur ne croit plus à la grande politique, aux utopies, aux unions internationales et à la mondialisation des capitaux. Aucun de ces idéaux ne saurait être à lui seul salvateur pour l'humanité. Il faut surmonter ces tentations et élaborer une sorte de démocratie sans État (p. 97). Au sein de cette nouvelle politique, le geste d'un militant libre aura, suivant une expression de Malarmé, la grandeur d'une action restreinte (p. 118).

La politique sans parti ne pourra se réaliser qu'à la faveur d'un détour nécessaire et encore impensable par les véritables œuvres révolutionnaires que furent celles de Lénine et de Mao Tsé-toung (p. 152). Si l'idée d'une politique sans parti contenue en germe dans ces œuvres nous est encore impensable, c'est parce que l'esprit démocratique est aujourd'hui trop directeur, voire despotique, lui qui gère le FMI, hélas seul juge du progrès, et qui défend les droits généraux d'un homme privé de sa singularité. Le retour aux révolutionnaires de la politique devrait en outre permettre à la philosophie de se libérer des attaches d'une idéologie contraignante. Car il ne fait nul doute que toute philosophie entretient des rapports plus ou moins avoués mais toujours concrets avec la politique (écoles, chefs, réunions, propagation de la nouvelle et règlements de compte). Elle représente même un danger lorsqu'elle se constitue en État. C'est alors qu'elle devient défavorable à la polyvalence de la pensée et qu'elle ne prend plus aucune décision en conformité avec les besoins singuliers de son milieu.

Badiou identifie avec beaucoup de justesse certains non-dits des philosophies contemporaines et détermine de façon fort stimulante les motifs d'une philosophie novatrice. L'audace calculée de son système appelle bien sûr un certain nombre de questions. Nous en proposons trois, relatives à chacune des procédures de vérité étudiées. D'abord, tel que le demandait pertinemment F. Wahl dans sa préface à *Conditions*, la volonté de dénouer la liaison entre la philosophie et la pensée ne fait-elle pas apparaître une suture plus profonde, voire assujettissante, entre la philosophie et la science mathématique? Ensuite, si le poème n'a pas réussi à conquérir sa propre laïcisation, d'autres formes d'expressions comme la musique ou la peinture ne sont-elles pas parvenues à dépasser l'option messianique en art? Et enfin, quelle différence établir entre, d'une part, les décisions de militants qui n'adhèrent à aucun parti politique et pour qui masse et État sont déliées et, d'autre part, un régime anarchique qui laisse libre cours à tous les désirs?

ALAIN BEAULIEU
Université de Paris VIII

Mathieu Kessler, *L'esthétique de Nietzsche*. Paris,
PUF, (coll. « Thémis Philosophie »), 1998, 259 p.

« Seule esthétique générale de Nietzsche publiée en français », l'œuvre de Mathieu Kessler se veut une présentation juste de l'esthétique de Nietzsche et non une simplification comme on en retrouve partout dans la littérature. Il s'agit, pour Kessler, de remettre les pendules à l'heure et de discuter des véritables enjeux de l'esthétique

Mathieu Kessler, *L'esthétique de Nietzsche*, Paris, PUF, (coll. « Thémis Philosophie »), 1998, 259 p.

Christine Daigle

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004913ar>
DOI : <https://doi.org/10.7202/004913ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Daigle, C. (2000). Compte rendu de [Mathieu Kessler, *L'esthétique de Nietzsche*, Paris, PUF, (coll. « Thémis Philosophie »), 1998, 259 p.] *Philosophiques*, 27 (1), 211–215. <https://doi.org/10.7202/004913ar>

politique) à cette différence près, fort appréciable cependant, que la liberté de penser du militant sans parti n'est jamais accordée au préalable au souverain. Le militant sans parti, qui devient le plus apte à prendre des décisions politiques, est donc celui qui ne prête allégeance à aucun souverain et à aucun État. Il n'attend plus rien des grands mouvements révolutionnaires et se connecte sur une multiplicité d'idéologies sans jamais en considérer une seule comme parfaitement achevée. L'auteur ne croit plus à la grande politique, aux utopies, aux unions internationales et à la mondialisation des capitaux. Aucun de ces idéaux ne saurait être à lui seul salvateur pour l'humanité. Il faut surmonter ces tentations et élaborer une sorte de démocratie sans État (p. 97). Au sein de cette nouvelle politique, le geste d'un militant libre aura, suivant une expression de Malarmé, la grandeur d'une action restreinte (p. 118).

La politique sans parti ne pourra se réaliser qu'à la faveur d'un détour nécessaire et encore impensable par les véritables œuvres révolutionnaires que furent celles de Lénine et de Mao Tsé-toung (p. 152). Si l'idée d'une politique sans parti contenue en germe dans ces œuvres nous est encore impensable, c'est parce que l'esprit démocratique est aujourd'hui trop directeur, voire despotique, lui qui gère le FMI, hélas seul juge du progrès, et qui défend les droits généraux d'un homme privé de sa singularité. Le retour aux révolutionnaires de la politique devrait en outre permettre à la philosophie de se libérer des attaches d'une idéologie contraignante. Car il ne fait nul doute que toute philosophie entretient des rapports plus ou moins avoués mais toujours concrets avec la politique (écoles, chefs, réunions, propagation de la nouvelle et règlements de compte). Elle représente même un danger lorsqu'elle se constitue en État. C'est alors qu'elle devient défavorable à la polyvalence de la pensée et qu'elle ne prend plus aucune décision en conformité avec les besoins singuliers de son milieu.

Badiou identifie avec beaucoup de justesse certains non-dits des philosophies contemporaines et détermine de façon fort stimulante les motifs d'une philosophie novatrice. L'audace calculée de son système appelle bien sûr un certain nombre de questions. Nous en proposons trois, relatives à chacune des procédures de vérité étudiées. D'abord, tel que le demandait pertinemment F. Wahl dans sa préface à *Conditions*, la volonté de dénouer la liaison entre la philosophie et la pensée ne fait-elle pas apparaître une suture plus profonde, voire assujettissante, entre la philosophie et la science mathématique? Ensuite, si le poème n'a pas réussi à conquérir sa propre laïcisation, d'autres formes d'expressions comme la musique ou la peinture ne sont-elles pas parvenues à dépasser l'option messianique en art? Et enfin, quelle différence établir entre, d'une part, les décisions de militants qui n'adhèrent à aucun parti politique et pour qui masse et État sont déliées et, d'autre part, un régime anarchique qui laisse libre cours à tous les désirs?

ALAIN BEAULIEU
Université de Paris VIII

Mathieu Kessler, *L'esthétique de Nietzsche*. Paris,
PUF, (coll. « Thémis Philosophie »), 1998, 259 p.

« Seule esthétique générale de Nietzsche publiée en français », l'œuvre de Mathieu Kessler se veut une présentation juste de l'esthétique de Nietzsche et non une simplification comme on en retrouve partout dans la littérature. Il s'agit, pour Kessler, de remettre les pendules à l'heure et de discuter des véritables enjeux de l'esthétique

nietzschéenne. Il souhaite ainsi éclairer ceux qui ont « acquis la fâcheuse tendance à se tenir quitte de la philosophie de l'art de Nietzsche en se limitant à un simple exposé des thèses défendues par *La naissance de la tragédie* » (p. 110).

La thèse de Kessler, qu'il qualifie de paradoxale, est que la rivalité apollinien/dionysien se dissout à mesure que l'esthétique de Nietzsche évolue, pour se conclure non pas par une victoire de Dionysos mais bien plutôt par une soumission de Dionysos à l'influence d'Apollon. Contrairement à l'interprétation courante qui veut que l'esthétique finale de Nietzsche (de même que l'ensemble de sa philosophie) soit purement dionysienne, Kessler souhaite montrer qu'« Apollon devient l'*éminence grise* de Dionysos » (p. 13). Kessler parle d'une « réforme théorique » de la figure de Dionysos qui vient à se transformer à un point tel que l'on peut parler d'un premier et d'un second Dionysos. C'est que, selon lui, l'esthétique nietzschéenne subit une évolution très importante. Son propos sera donc de retracer les étapes et éléments faisant partie de cette évolution qui nous mène d'une « métaphysique d'artiste », soit une esthétique qualifiée de romantique, à une « physiologie de l'art », soit une esthétique dite classique. Toutefois, Kessler prendra bien soin de montrer que cela ne résulte pas en une contradiction fondamentale mais plutôt dans un changement majeur de perspective.

Sa présentation de l'« esthétique négative » de Nietzsche démontre que celle-ci « est partiellement recouverte par un mythe qui accorde à la conception d'une sorte de "formalisme dionysiaque" beaucoup plus d'importance qu'il n'en possède dans les textes eux-mêmes » (p. 35). Se posant contre des esthéticiens nietzschéens tels Luc Ferry et Michel Onfray, Kessler signale qu'en fait on doit parler d'« unité organique de l'art total » lorsqu'on parle de la première esthétique de Nietzsche. Il souligne que « *La naissance de la tragédie* insiste sur la nécessité et l'intérêt esthétique de conserver l'art apollinien et l'art dionysiaque dans une étroite union » (p. 36). Mais ces notions d'art total et d'esthétique négative évolueront elles aussi et ne constituent qu'un moment de l'esthétique nietzschéenne.

Kessler fait des remarques très intéressantes sur le goût de Nietzsche et son lien avec l'esthétique. D'une part, Nietzsche affiche un goût qui est en réaction contre l'esthétique de Schopenhauer et de Wagner et, d'autre part, « il crée dans l'esprit de son temps » (p. 42). Kessler remarque que Nietzsche effectue une « rationalisation de son goût pour des raisons non esthétiques (p. 43) ». Mais, en fait, la tendance naturelle de Nietzsche l'entraîne dans le romantisme et c'est sa culture philosophique et esthétique qui l'amène à ce que Kessler appelle un quasi-militantisme pour le classicisme. Ce tournant vers le classicisme s'accompagne, selon Kessler, d'un souci d'approfondissement philosophique des problèmes moraux initialement traités par la tragédie grecque. Il dit ainsi que « la philosophie de la *cruauté* et de l'*amor fati* prendra donc le relais de l'*esthétique de la dissonance* » (p. 51-52).

Le prochain objet d'analyse de Kessler est ce qu'il appelle l'esthétique du rire et des larmes. Il signale d'abord que l'opposition Apollon/Dionysos ou encore celle entre Dionysos et le Christ en camoufle une autre beaucoup plus fondamentale, soit l'opposition entre une esthétique et une philosophie positive ou négative. Or, comme le souligne bien Kessler, la nouvelle attitude dionysiaque qui se manifeste dans l'*Essai d'autocritique* est une attitude essentiellement positive, affirmative et créatrice de valeurs. Dionysos se transforme : de figure tentant de dissuader l'homme de vivre, il devient un réformateur de l'individu en tentant de le rendre plus méchant et, donc, plus naturel. Kessler remarque que Dionysos revêt de plus en plus les caractéristiques du surhomme, mais ce n'est pas tout. Le second Dionysos « incarne le monde de l'appar-

rence en représentant l'Éternel retour et le "sens de la Terre", il est nécessaire qu'il revête certaines des caractéristiques naguère propres à Apollon » (p. 72). Kessler montre bien que l'esthétique ne résiste pas à l'évolution que l'on retrouve par ailleurs dans la pensée de Nietzsche.

Malgré ce qui semble un retour en force de Dionysos, Kessler souligne que « l'esthétique de Nietzsche reculera de plus en plus devant ce qui est purement dionysiaque » (p. 82). La raison en est, selon Kessler, que Dionysos est une divinité beaucoup plus philosophique qu'Apollon. Kessler pense donc que l'art dans sa fonction générale correspond à une logique apollinienne alors que la philosophie correspondrait à une logique dionysiaque. L'art dionysiaque est celui qui est « engagé par la philosophie à penser le sens tragique de la vie » (d'où l'importance majeure de Dionysos dans *La naissance de la tragédie*). Kessler souligne que « la finalité du monde de l'art dionysiaque est le dévoilement de la *vérité* du sens tragique de l'existence, ce qui a toujours été *en même temps* la définition fondamentale et invariable de la (*bonne*) philosophie d'après Nietzsche » (p. 89).

Il est à remarquer que cette transformation de Dionysos s'accompagne d'une disparition : Dionysos est absent et pas un mot n'est dit à son propos de 1872 à 1885. C'est un fragment posthume qui rompt le silence à propos de Dionysos. Il est alors représenté comme figure unificatrice du dualisme métaphysique présent dans *La naissance de la tragédie*. Kessler remarque que la période « rationaliste » de Nietzsche est suivie d'un retour à l'« âge tragique » de sa jeunesse qui entraîne une « mise à jour » de la figure de Dionysos. Ce

philosophe-artiste peut se caractériser par le fait qu'il ne se contente plus de *penser* le monde ni d'en *transformer* un simple fragment, mais qu'il a pour ambition suprême et inégalée de veiller à la réorganisation du monde en sa totalité. [...] Ses fonctions sont d'abord caractérisées par une volonté de modifier les principes métaphysiques et moraux du judéo-christianisme, en proposant le retour à une sorte de religion plus proche de la nature sauvage et contradictoire de l'homme. Enfin, Dionysos remplira la fonction insigne d'*Antithèse du type de l'homme religieux chrétien*, il sera, littéralement cette fois, « l'antéchrist » (p. 94-95).

Kessler conclut la première partie de son livre en discutant du déclin final de l'antagonisme radical Apollon/Dionysos, soit de l'absorption d'Apollon par Dionysos. En tant qu'incarnation du grand « oui », Dionysos est également incarnation de l'antagonisme. Selon Kessler, « la vision que Nietzsche se fait de *La naissance de la tragédie* est donc infléchie par l'évolution de Dionysos et son identification avec la conquête beaucoup plus tardive d'une solution entièrement philosophique au problème du pessimisme (*l'amor fati*) » (p. 150), solution que ne peut apporter Apollon mais que Dionysos, en tant qu'incarnation de l'antagonisme, peut apporter.

La seconde partie de l'ouvrage de Kessler se consacre à une analyse du formalisme classique de Nietzsche. Il souligne d'abord que « le classicisme de Nietzsche se caractérise par une grande indifférence au contenu tandis qu'une bonne partie des controverses esthétiques sur l'art classique proviennent de son respect du principe de l'*imitation* et d'une relation qualifiée avec la *vérité* » (p. 155). Par ailleurs, et aussi surprenant que cela puisse paraître, le classicisme de Nietzsche se présente comme un « prolongement et une radicalisation d'un processus de pensée déjà entrepris par Kant » (p. 157). Le rejet kantien de tout lien entre le jugement de goût et les jugements

logiques ou de connaissance se réaffirme chez Nietzsche, ce qui constitue d'ailleurs le point principal « interdisant une assimilation de son esthétique avec la conception classique d'un art au service du vrai » (p. 157). Kessler remarque qu'il va même à l'inverse de la conception classique traditionnelle en affirmant que la vérité est laide et que la tromperie seule est belle. Nietzsche présente donc un classicisme tout à fait particulier et distinct de toute théorie esthétique. S'ajoute à cela le principe que l'art est un remède opposé au nihilisme, en quoi l'on peut voir en Schopenhauer son adversaire en termes de philosophie de l'art.

Kessler parle d'un classicisme de la forme chez Nietzsche. Voici son explication :

La forme de l'art classique — le style classique — c'est donc le « contenu » du classicisme, selon Nietzsche, tandis que son contenu, à savoir l'idéologie classique, le message classique d'inspiration platonicienne, n'est que la « forme » particulière que revêt un style déterminé à une étape historique donnée. La forme de tout classicisme est une nécessité tandis que son contenu n'est qu'accidentel. [...] Seule la notion de style est déterminante aux yeux de Nietzsche, car l'idéologie n'est qu'un masque et une cote toujours mal taillée pour le génie artistique. [...] on déterminera le classicisme de Nietzsche comme « corps du classicisme », c'est-à-dire utilisation du style classique sans l'argument de l'esprit classique (p. 166-167).

Ce que Kessler souhaite montrer, c'est que la notion de « formalisme classique » est la plus adéquate pour désigner la nouvelle fonction « médicale » de l'esthétique de Nietzsche qui se veut physiologie de l'art. Il s'agit, pour Nietzsche, de se guérir du romantisme des premiers pas esthétiques.

Dans un intéressant chapitre sur l'esthétique musicale, Kessler montre que Nietzsche inverse sa thèse sur la musique depuis *La naissance de la tragédie* : « cette modification destitue Dionysos, essentiellement patron de l'art musical, de sa position de prééminence, par ailleurs de plus en plus affirmée par rapport à Apollon » (p. 179). Kessler illustre bien cette modification des visées de Nietzsche par les modifications qu'a subi *La naissance de la tragédie* à travers les trois différentes éditions qu'elle a eues du vivant de Nietzsche. Il vient à juger la musique avec d'autres critères et son esthétique se transforme : « Elle conteste l'indépendance et l'autonomie de la musique par rapport aux critères esthétiques inspirés par d'autres arts » (p. 198). « La musique perd son statut d'art médiateur de l'Unité originaire exprimée à travers le temps. Nietzsche la caractérise maintenant comme d'un goût moins sûr et moins durable que celui des inventions des arts plastiques ou du monde des idées » (p. 202). L'objectif de Nietzsche à travers toutes ces modifications est d'abattre le mysticisme auquel collaborait sa première esthétique musicale. Concernant l'évolution de cette esthétique, on pourra dire avec Kessler qu'« abandonnant sa métaphysique initiale, elle récuse la représentation du devenir que la rythmique de la poésie et de la musique exprime plus adéquatement que les arts plastiques pour adorer le mythe de l'instant pétrifié et spatialisé dans l'éternité factice de la belle apparence apollinienne » (p. 207).

Le dernier chapitre s'intéresse à la logique de la physiologie de l'art et tente de déterminer la fonction de l'art dans l'esthétique nietzschéenne. Dans ce chapitre, Kessler souligne que la doctrine de « l'art pour l'art » est la plus éloignée de l'esthétique de Nietzsche. La seule chose que les deux puissent avoir en commun est le goût pour la pure forme. Quelle est donc la fonction de l'art, après l'abandon de toute fonction métaphysique?

« Être trompé d'inoffensive manière », telle est la fonction homéopathique de l'erreur inculquée à petites doses par l'art. En effet, ni l'erreur absolue ni une dose massive de vérité ne sont souhaitables eu égard à la vie. [...] Cette fonction particulière de *distillation homéopathique de l'erreur* attribuée à la *beauté* est une constante de l'esthétique de Nietzsche (p. 229-230).

Et quel est le lien final entre art et philosophie? « L'art doit représenter une interprétation positive du monde où surtout le chaos est maîtrisé, tandis que la philosophie doit se préserver exclusivement la tâche de "voir clair dans ce qui est" en dehors de toute interprétation humaine et selon la méthode fatalement nihiliste de la probité philologique » (p. 234). Après 1876, l'art et la philosophie seront radicalement séparés. Kessler souligne que malgré que cette séparation radicale semble mal s'accorder avec la pluralité des entreprises marginales de Nietzsche, « il n'en demeure pas moins que cette rigueur esthétique est le meilleur garant du formalisme classique de Nietzsche dans la mesure où celui-ci se caractérise de façon cohérente comme refusant tout rapport avec un *contenu* quel qu'il soit » (p. 235). En fin de compte, ce qui, selon Kessler, légitimise la physiologie de l'art c'est qu'une philosophie pessimiste a besoin d'un art optimiste. Au point de vue philosophique, c'est l'*amor fati* qui solutionne l'impasse du nihilisme, « mais il ne semble pas, selon Nietzsche, à la portée des moyens d'expression artistiques qui ne peuvent affirmer la vie qu'à travers la belle illusion » (p. 237).

Kessler conclut son étude en soulignant que Nietzsche prend ses distances vis-à-vis une justification purement esthétique de l'existence, telle qu'exprimée dans *La naissance de la tragédie*, notion trop restrictive en comparaison avec celle d'*amor fati*. Mais, alors que « la philosophie est capable de *penser* jusqu'au bout le nihilisme comme l'unique loi du développement de l'histoire de l'humanité et donc des civilisations, seuls l'art et la grande politique sont susceptibles de *transformer* un monde devant rester dans l'ignorance des raisons qui le gouvernent » (p. 244). Mais entre la politique et l'art, l'art est ce en quoi Nietzsche place sa confiance. Kessler conclut donc :

L'art se trouve donc placé au cœur de la problématique de la civilisation chez Nietzsche, car lui seul permet de cultiver les âmes en tant qu'elles sont incarnées dans des corps, tandis qu'un dressage purement moral et prescriptif ne peut aboutir qu'à la répression de tous les instincts affirmatifs, en les délimitant par des règles universelles aux conséquences négatrices et donc nihilistes (p. 248).

L'étude de Kessler remplit ses objectifs initiaux, soit de présenter une analyse juste et complète de l'esthétique de Nietzsche. Son analyse, bien supportée par les documents (je souligne l'utilisation très judicieuse qu'il fait des fragments posthumes), montre bien en quoi nombre d'interprétations de l'esthétique nietzschéenne sont en fait unilatérales et mettent de côté tout un pan de son esthétique qui est essentiel à toute interprétation adéquate. L'entreprise de Kessler s'inscrit donc dans le courant de réinterprétation de l'œuvre nietzschéenne des dernières années et accomplit la réinterprétation de l'esthétique qui manquait dans la littérature.

CHRISTINE DAIGLE
Université de Montréal
Université Concordia

Bernard Lamy, *La Rhétorique ou l'art de parler* (éd. critique établie par Benoît Timmermans), Paris, PUF, (coll. « L'interrogation philosophique »), 1998, 598 p.

Lucie Desjardins

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004914ar>
DOI : <https://doi.org/10.7202/004914ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)
1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Desjardins, L. (2000). Compte rendu de [Bernard Lamy, *La Rhétorique ou l'art de parler* (éd. critique établie par Benoît Timmermans), Paris, PUF, (coll. « L'interrogation philosophique »), 1998, 598 p.] *Philosophiques*, 27 (1), 216–219.
<https://doi.org/10.7202/004914ar>

Bernard Lamy, *La Rhétorique ou l'art de parler* (éd. critique établie par Benoît Timmermans). Paris, PUF, (coll. « L'interrogation philosophique »), 1998, 598 p.

L'édition de *La Rhétorique ou l'art de parler* de Bernard Lamy que Benoît Timmermans a procurée au public en 1998 relève du tour de force. Paru pour la première fois en 1675 sous le titre *De l'art de parler*, l'ouvrage sera maintes fois remanié et amplifié par son auteur à l'occasion des 29 éditions recensées dans lesquelles on passe progressivement de 288 pages à 444 pages. En présentant, de façon habile, le texte avec ses nombreuses variations, Timmermans fait plus que proposer les différentes strates du travail d'écriture. L'intérêt de ces multiples remaniements, écrit-il, est « qu'ils fournissent une suite de clichés représentatifs des états successifs de la problématique rhétorique, mais aussi philosophiques, qui animent les disciplines centrées sur le langage à la charnière des XVII^e et XVIII^e siècles » (p. 5).

L'ouvrage de Lamy entend se distinguer radicalement d'avec les rhétoriques que l'on retrouve habituellement à la même époque. Dès la préface, l'auteur les dénonce :

Les rhétoriques ordinaires sont faites de manière qu'elles sont presque entièrement inutiles. [...] Aussi il y a bien des gens qui font peu de cas de la rhétorique, ce qui n'arriverait pas si elle était bien traitée, c'est-à-dire si elle donnait effectivement la connaissance de ce qu'elle devrait enseigner ; ce qu'il est bon de considérer ici (p. 14).

En contestant l'utilité des « rhétoriques ordinaires », Lamy effectue une intrusion remarquée dans le genre codifié du traité oratoire. Une tradition, déjà deux fois millénaire, voulait que le manuel de rhétorique se livrât à l'énumération des genres et des diverses parties de l'art oratoire avant de s'étendre sur certaines d'entre elles, en général sur l'*inventio*, puis, à un degré moindre, sur les deux suivantes, la *dispositio* et l'*elocutio*. C'est ce que fait, par exemple, René Bary dans sa *Rhétorique françoise* (1659) : centrée sur l'*inventio*, celle-ci fournit à profusion des listes de lieux généraux, spécifiques, extrinsèques et intrinsèques. L'élève apprend alors à discourir comme en dialectique, en égrenant définition, énumération des parties, etc.

On ne retrouve pas, chez l'oratorien Lamy, de listes de lieux logiques : ceux-ci sont « dangereux pour les personnes qui ont peu de savoir, parce qu'elles se contentent de ces preuves qui se trouvent facilement, et qu'elles ne prennent pas la peine d'en chercher d'autres qui soient plus solides » (p. 453). L'art de parler ne sera donc qu'une partie de l'art de persuader, celle qui, dans l'ancienne rhétorique, avait pour nom *elocutio*. Et si on peut encore reconnaître bon nombre de notions empruntées à la rhétorique classique, c'est suivant un dispositif bien différent et avec des explications inédites.

Le premier livre traite de grammaire et de logique. À la suite de Port-Royal, Lamy considère que « les paroles sont des signes qui représentent les choses qui se passent dans notre esprit », si bien qu'elles sont comme une « peinture de nos pensées, que la langue est le pinceau qui trace cette peinture et que les mots dont le discours est orné en sont les couleurs » (p. 35). La multitude des choses, le nombre infini de manières dont l'esprit peut les envisager exige toutefois un lexique lui-même infini. Dans ces conditions, il est impossible d'imaginer une langue qui réponde parfaitement aux besoins de la conceptualisation en établissant une correspondance univoque entre les mots et les choses. Aussi le deuxième livre est-il entièrement consacré aux tropes et aux figures. Toutefois, il offre bien plus que cette sorte de répertoire des

tropes et des figures illustré d'exemples que les rhétoriques se sont plu, au fil des siècles, à dresser avec une précision toujours accrue. Lamy tente plutôt d'exposer les raisons de l'apparition des tropes et des figures dans le discours. Il en arrive alors à une question essentielle : celle des passions. Selon lui, le langage naturel est traversé par la passion, car il véhicule des idées par des mots qui ne traduisent pas forcément des idées claires et distinctes. En effet, le discours d'un sujet en proie à une ou plusieurs passions se distingue du discours de celui qui « parle sans émotion » :

Ainsi les paroles répondant à nos pensées, le discours d'un homme qui est ému ne peut être égal. Quelquefois il est diffus, et il fait une peinture exacte des choses qui sont l'objet de sa passion : il dit la même chose en cent façons différentes. Une autre fois son discours est coupé, les expressions en sont tronquées ; cent choses y sont dites à la fois : il est entrecoupé d'interrogations, d'exclamations ; il est interrompu par de fréquentes digressions ; il est diversifié par une infinité de tours particuliers et de manières de parler différentes. Ces tours et ces manières de parler sont tout aussi faciles à distinguer d'avec les façons de parler ordinaires que les traits d'un visage irrité d'avec ceux d'un visage calme (p. 182).

Dès lors, ces « tours particuliers » et ces « manières de parler différentes » ne sont que les signes d'une pensée bousculée par une expérience perceptive modifiée ou altérée par les passions. Aussi le langage de la passion n'est-il « pas libre », mais résulte d'un mouvement naturel, si bien que ce n'est plus le concept de signe, mais celui de figure qui servira désormais à se mouvoir dans cet univers. En effet,

les passions font que l'on considère les choses d'une autre manière que l'on ne le fait dans le repos et dans le calme de l'âme : elles grossissent les objets, elles y attachent l'esprit ; et font presque autant d'impression sur lui que les choses mêmes (p. 181).

C'est parce que les passions font que l'on considère les choses d'une autre manière qu'on ne le fait dans le « repos et dans le calme de l'âme » qu'elles modifient le rapport vécu au monde et, par conséquent, le discours. Reste à établir les modalités de ces modifications. Tout concourt alors à faire en sorte que la langue ne semble plus donner seulement accès à la pensée, mais à un regard particulier sur le monde que modulent et orientent les mouvements des passions.

Dans cette perspective, il ne saurait exister de passion sans une forme de représentation facilement reconnaissable et c'est dans des termes qui rappellent *La conférence sur l'expression générale et particulière* de Charles Le Brun (1668) que Lamy observe tout aussi bien que le premier effet des passions tient à une altération qu'elles provoquent autant sur un visage vierge de toute marque d'agitation, que dans le discours dépourvu de figures d'un homme qu'aucune passion n'agite.

Comme on lit sur le visage d'un homme ce qui se passe dans son cœur ; que le feu de ses yeux, les rides de son front, le changement de couleur de son visage, sont les marques évidentes des mouvements extraordinaires de son âme ; les tours particuliers de son discours ; les manières de s'exprimer éloignées de celles que l'on garde dans la tranquillité, sont les signes et les caractères des agitations, dont son esprit est ému dans le temps qu'il parle (p. 181).

À partir de ce parallèle entre figures du discours et marques des passions qui se peignent naturellement sur le visage, Lamy aborde ensuite la question de la représen-

tation des passions dans le discours. Plus que sur ces « manières de parler », l'attention se porte alors sur les raisons de ces manières. Reste à établir les modalités du rapport entre ces figures du discours et les passions qu'elles représentent. S'il y a une correspondance entre figures du discours et passions de l'âme, ce rapport semble difficilement intégrable dans un code. Les figures qui sont les caractères discursifs des passions viennent rendre compte d'un regard particulier sur les objets du monde, regard particulier qui sert, à son tour, d'argument.

Dès lors, l'essentiel de la figure ne réside plus dans l'écart, la déviation de langage par laquelle elle se réalise, mais dans l'état affectif qu'elle traduit et qu'elle vise à communiquer. Encore une fois, Lamy excède la rhétorique classique qui s'intéressait à l'*elocutio* en tant que technique ornementale et voyait dans « les fleurs de la rhétorique » un supplément destiné à agrémenter des œuvres dont la beauté devait d'abord consister dans la justesse d'ensemble. Parce qu'elles représentent les passions, les figures ne sont pas qu'un ornement du discours : elles sont « utiles et nécessaires » (p. 185). L'orateur peut s'en servir pour faire la peinture exacte d'une passion, pour convaincre qu'il est lui-même véritablement ému ou encore pour susciter une passion chez l'auditeur. Car les figures ne sont ici rien d'autre que la manière dont une passion modifie la perception et, par conséquent, le discours : elles sont un lieu qui permet à la passion de se trouver un langage propre et des signes authentiques à même de révéler une pensée conçue comme indissociable de la perception des choses. Aussi devront-elles s'appliquer à rendre le mouvement diffus de la pensée d'un individu livré à ses passions. En tenant compte de cette composante psychologique d'une perception des choses modifiée par la passion, Lamy s'emploiera, dans la suite du texte, à définir un certain nombre de figures susceptibles d'exprimer, mieux que d'autres, ce mouvement : l'exclamation, le doute, l'épanorthose, l'ellipse, l'aposiopèse, l'hyperbate, la paralipse, la répétition, la paronomase, le pléonasmе, le synonyme, l'hypotypose, la description, la distribution, l'antithèse, la similitude, la comparaison, la suspension, la prosopopée, la sentence, l'épiphonème, l'interrogation, l'apostrophe, l'épistrophe, la prolepse, la communication, la confession, le consentement et la périphrase (p. 190-217). On remarquera que ces figures conviennent aussi pour des discours qui ne mettent pas en cause une passion.

En ce sens, le regard particulier que suppose la passion vient s'opposer, chez Lamy, à la raison ou, pour mieux dire, à un rationalisme philosophique qui, au XVII^e siècle, vise à donner du monde une représentation qui se veut universelle. Les idées claires et distinctes de Descartes conduisent à une représentation univoque du monde, car rien de ce qu'on peut connaître par la raison dans le monde extérieur ne peut être différent de ce que représentent ces idées claires et distinctes, elles-mêmes garanties par un modèle logique qui en appelle, de manière ultime, à Dieu. Or, la question de la représentation des passions permet à Lamy d'introduire au cœur même de sa rhétorique, pourtant d'inspiration cartésienne, une théorie susceptible d'envisager la multiplicité des points de vue. Ce serait en effet se méprendre de croire que, si les figures marquent la passion, elles excluent toute forme de jugement : au contraire,

Le ton de la voix, et le tour qu'on prend fait connaître de quelle manière on regarde les choses dont on parle, le jugement qu'on en fait, et les mouvements dont on est animé à leur égard (p. 157).

Du moment que la passion nous fait considérer un objet, celui-ci perd son unicité et devient multiple. Les figures ne sont donc pas réglées sur une représentation

du monde à vocation universelle, mais bien sur un regard et une perception particulière indissociable du mouvement de la passion.

Le troisième livre s'intéresse à l'autre versant du langage, à la substance phonique, c'est-à-dire à ce que Lamy, dans la métaphore qu'il emprunte à Descartes, appelle « le corps du discours ». Cette corporéité doit être analysée dans toute son étendue et avec le secours de l'ensemble des disciplines susceptible de donner à comprendre cette corporéité. Ce livre va donc inclure, à côté des notions de phonétique générale, un vaste ensemble de questions relatives aux sciences, à la philosophie et à la poétique qui conduira l'auteur, dans le dernier livre, à aborder la question du style. Comme le discours est le « caractère de l'âme », il n'est pas étonnant qu'il y ait autant de différents styles qu'il y a de personnes qui parlent ou écrivent. De même, les hommes qui vivent dans des zones climatiques différentes ont un style différent. Ici, la théorie des climats vise à rendre compte de la « couleur particulière » propre au rendu discursif de la perception. Au surplus, si Lamy reprend la tripartition reçue des styles (style sublime, style moyen et style simple), ce ne sera pas sans y ménager un espace à la passion :

lorsque les choses sont grandes, et que l'on ne peut les envisager sans ressentir quelque grand mouvement, le style qui les décrit doit être nécessairement animé, plein de mouvement, enrichi de Figures, de toutes sortes de Métaphores. Si le sujet que l'on traite n'a rien d'extraordinaire, si on le peut considérer sans être touché de passion ; le style doit être simple (p. 378).

Négligée par les études sur la rhétorique qui ont longtemps considéré que la rhétorique française naissait avec Dumarsais et Fontanier, la *Rhétorique* de Lamy est tenue aujourd'hui, à juste titre, pour un texte essentiel lorsqu'on s'intéresse à l'histoire des idées en général et à l'histoire de la rhétorique en particulier. Essentiel, car l'attention que Lamy porte à la question des passions dans le discours force à reconsidérer l'idée, trop souvent répandue depuis Michel Foucault, que le ^{xvii}^e siècle tout entier est habité par le seul désir d'une langue qui viendrait refléter l'univocité des signes. Essentiel, parce que Lamy participe à un ample mouvement qui ira s'intensifiant au cours du siècle suivant et qui cherchera, par-delà l'enseignement de la rhétorique, à placer les phénomènes sensoriels et psychologiques au centre de la réflexion sur le langage et de la problématique esthétique. L'excellente édition de Benoît Timmermans fait plus que remettre au jour ce texte capital : elle présente surtout l'avantage de retracer les strates successives du travail de Lamy et, par là, permet de mieux comprendre le passage du « Grand Siècle » au « Siècle des Lumières ».

LUCIE DESJARDINS

Département d'études littéraires
Université du Québec à Montréal

Jocelyn Benoist, Robert Brisart, Jacques English, *Limitaires phénoménologiques : recherches sur le développement de la théorie de la signification de Husserl*. Bruxelles, Publications des Facultés universitaires Saint-Louis, 1998, 281 p.

Cet ouvrage, dont le thème demeure encore inusité dans la littérature de langue française, tente d'investiguer à nouveaux frais la théorie de la signification de Husserl. En

Jocelyn Benoist, Robert Brisart, Jacques English, *Liminaires phénoménologiques : recherches sur le développement de la théorie de la signification de Husserl*, Bruxelles, Publications des Facultés universitaires Saint-Louis, 1998, 281 p.

Guillaume Fréchette

Le matérialisme contemporain
Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004922ar>

DOI : <https://doi.org/10.7202/004922ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Fréchette, G. (2000). Compte rendu de [Jocelyn Benoist, Robert Brisart, Jacques English, *Liminaires phénoménologiques : recherches sur le développement de la théorie de la signification de Husserl*, Bruxelles, Publications des Facultés universitaires Saint-Louis, 1998, 281 p.] *Philosophiques*, 27 (1), 219–225.
<https://doi.org/10.7202/004922ar>

du monde à vocation universelle, mais bien sur un regard et une perception particulière indissociable du mouvement de la passion.

Le troisième livre s'intéresse à l'autre versant du langage, à la substance phonique, c'est-à-dire à ce que Lamy, dans la métaphore qu'il emprunte à Descartes, appelle « le corps du discours ». Cette corporéité doit être analysée dans toute son étendue et avec le secours de l'ensemble des disciplines susceptible de donner à comprendre cette corporéité. Ce livre va donc inclure, à côté des notions de phonétique générale, un vaste ensemble de questions relatives aux sciences, à la philosophie et à la poétique qui conduira l'auteur, dans le dernier livre, à aborder la question du style. Comme le discours est le « caractère de l'âme », il n'est pas étonnant qu'il y ait autant de différents styles qu'il y a de personnes qui parlent ou écrivent. De même, les hommes qui vivent dans des zones climatiques différentes ont un style différent. Ici, la théorie des climats vise à rendre compte de la « couleur particulière » propre au rendu discursif de la perception. Au surplus, si Lamy reprend la tripartition reçue des styles (style sublime, style moyen et style simple), ce ne sera pas sans y ménager un espace à la passion :

lorsque les choses sont grandes, et que l'on ne peut les envisager sans ressentir quelque grand mouvement, le style qui les décrit doit être nécessairement animé, plein de mouvement, enrichi de Figures, de toutes sortes de Métaphores. Si le sujet que l'on traite n'a rien d'extraordinaire, si on le peut considérer sans être touché de passion ; le style doit être simple (p. 378).

Négligée par les études sur la rhétorique qui ont longtemps considéré que la rhétorique française naissait avec Dumarsais et Fontanier, la *Rhétorique* de Lamy est tenue aujourd'hui, à juste titre, pour un texte essentiel lorsqu'on s'intéresse à l'histoire des idées en général et à l'histoire de la rhétorique en particulier. Essentiel, car l'attention que Lamy porte à la question des passions dans le discours force à reconsidérer l'idée, trop souvent répandue depuis Michel Foucault, que le ^{xvii}^e siècle tout entier est habité par le seul désir d'une langue qui viendrait refléter l'univocité des signes. Essentiel, parce que Lamy participe à un ample mouvement qui ira s'intensifiant au cours du siècle suivant et qui cherchera, par-delà l'enseignement de la rhétorique, à placer les phénomènes sensoriels et psychologiques au centre de la réflexion sur le langage et de la problématique esthétique. L'excellente édition de Benoît Timmermans fait plus que remettre au jour ce texte capital : elle présente surtout l'avantage de retracer les strates successives du travail de Lamy et, par là, permet de mieux comprendre le passage du « Grand Siècle » au « Siècle des Lumières ».

LUCIE DESJARDINS

Département d'études littéraires
Université du Québec à Montréal

Jocelyn Benoist, Robert Brisart, Jacques English, *Limitaires phénoménologiques : recherches sur le développement de la théorie de la signification de Husserl*. Bruxelles, Publications des Facultés universitaires Saint-Louis, 1998, 281 p.

Cet ouvrage, dont le thème demeure encore inusité dans la littérature de langue française, tente d'investiguer à nouveaux frais la théorie de la signification de Husserl. En

effet, sur la base des publications des dernières années de la collection des *Husserliana*, plusieurs efforts de relecture de la première partie de l'œuvre de Husserl ont été entrepris, auxquels il faut dire que les auteurs du présent ouvrage ont contribué dans une mesure importante chez les francophones. C'est donc essentiellement sur cette première partie de l'œuvre de Husserl qu'insistent les auteurs, celle que couvre la rédaction de la thèse d'habilitation de 1887 (*Über den Begriff der Zahl*) jusqu'aux premières ébauches du vocabulaire noématique en 1911/1912. Cette période voit paraître des ouvrages comme la *Philosophie de l'arithmétique* (1891), les *Recherches logiques* (1900/1901), ainsi que des leçons telle que la *Théorie de la signification* (1908), qui sont de première importance quant au thème de la signification, en ce qu'ils tentent à la fois de définir le statut de la signification à l'intérieur même de la relation intentionnelle, mais également de faire composer une analyse linguistique et une grammaire dans l'élaboration du langage d'une ontologie.

Dans « La découverte des processus signitifs dans la première œuvre de Husserl (1887-1891) » (p. 9-62), Brisart insiste sur le passage des processus abstractifs, issus de l'abstraction mathématique, aux processus signitifs, un passage qui manifesterait la première élaboration d'une théorie de la signification chez Husserl : « c'est par la découverte des processus signitifs intrinsèques à l'arithmétique et, avec elle, par la mise en évidence de la distinction cardinale entre signification et intuition, que s'est décidée pour la première fois la mise en perspective phénoménologique de l'objet intentionnel en tant qu'objet idéal » (p. 10). En d'autres termes, cette découverte des processus signitifs par Husserl est à considérer comme inaugurant celle qui l'amène à distinguer le contenu de la représentation : le passage à une conception symbolique de la représentation du nombre dans la *Philosophie de l'arithmétique* est motivé par l'insuffisance d'une définition de l'acte de dénombrer en tant que processus abstractif, insuffisance qui permet seulement une synthèse des concepts de nombre au niveau des *abstracta*. Aussi, une représentation du nombre de ce point de vue se composerait de deux parties psychologiques, soit le processus d'unification collective (synthèse donnant les *abstracta*) et le quelque chose. Cette distinction laisse toutefois indéterminée la question du contenu des représentations : dit autrement, les nombres sont représentés selon cette première conception comme du quelque chose, « au sens où quelque chose est une formation propre à un acte psychique qui se représente un contenu quel qu'il soit, abstraction faite de la teneur particulière de ce contenu » (p. 19). La représentation du nombre serait donc une représentation sans contenu ou, dit autrement, un acte qui se représente l'ensemble (qui lie de manière collective), ce qui expliquerait l'acte de dénombrer comme indépendant de ce qui se trouve dans le contenu de la représentation. Comme on le sait, cette question de la représentation est posée par Husserl sur un terrain prioritairement mathématique, cette thèse d'habilitation de 1887 visant en quelque sorte à expliciter l'intuition de Weierstrass selon laquelle un développement rigoureux de l'analyse supérieure doit provenir de l'arithmétique élémentaire. C'est là un point que l'auteur aurait pu gagner à expliciter davantage, car il est souvent de mise de considérer les premiers écrits de Husserl à la lumière de son intérêt pour la philosophie brentanienne, alors qu'on oublie parfois que cet intérêt était motivé par les questions de fondement des mathématiques qu'il développait sous l'influence de Weierstrass. Quant à la conception symbolique de la représentation du nombre, elle fait passer l'analyse des nombres du point de vue psychologique à un point de vue dit signitif ou symbolique, ce que l'auteur explique par le fait que les opérations numériques en arithmétique n'opèrent sur les nombres qu'en tant que signes

ou symboles de concepts, et que ce sont seulement ces signes qui sont liés dans le calcul. Bien loin de se laisser dériver des concepts de l'arithmétique élémentaire, les nombres irrationnels apparaissent plutôt comme relevant de concepts symboliques dont une justification et une fondation ultimes n'auraient donc pu simplement en appeler à une analyse du concept de nombre au sens le plus élémentaire. En ce sens, l'approche seulement symbolique doit être d'un genre différent de l'approche psychologique-descriptive. Par l'élaboration de cette distinction, l'auteur souhaite insister sur le fait que c'est par une réflexion sur les processus signitifs et leur rôle dans la construction de la représentation que Husserl se détache pour la première fois et de manière significative d'une tendance psychologiste à la Brentano. Aussi, la *Philosophie de l'arithmétique* ne serait pas si brentanienne si nous acceptons avec l'auteur qu'à partir de 1890, Husserl se détache progressivement d'une conception pratique de la logique, comme *Kunstlehre*, et qu'en ce sens, il n'aurait jamais pu être pleinement psychologiste car c'est d'abord eu égard à la question de la discipline fondatrice (arithmétique ou logique) que la psychologie aurait été avancée. Mais dans la mesure où la conclusion de la *Philosophie de l'arithmétique* suggère une dépendance de l'arithmétique à la logique, la question du psychologisme de la *Philosophie de l'arithmétique* n'est pas centrale, mais c'est plutôt celle de la critique de l'algébrisation de la logique (telle que proposée par Schröder à l'époque) qui serait déterminante. Or, selon l'auteur, la critique de l'algébrisation schröderienne de la logique, telle qu'elle fera l'objet d'un compte-rendu de Husserl en 1890 et sur lequel l'auteur nous promet un exposé paru récemment dans la revue *Recherches husserliennes*, serait déjà une critique de la logique comprise comme *Kunstlehre*. Partant de ce fait, il est difficile de concevoir quelle thèse est véritablement défendue dans la *Philosophie de l'arithmétique* : si l'ouvrage donne un rôle prioritaire à la logique sur l'arithmétique, c'est dans la mesure où cette logique est une science pratique dont les tâches sont prescriptives et non normatives. Si cette conception contrevient à l'algébrisation de la logique, cette contravention ne signifie pas pour autant que Husserl accorde à la logique un rôle de science théorique aux lois aprioriques. Dans les faits, la *Philosophie de l'arithmétique* marque peut-être la fin de l'étude psychologique de l'origine des nombres, non pas parce que les moyens d'investigation ne sont pas efficaces, mais parce que l'objet d'étude n'est plus considéré de façon aussi fondamentale qu'il l'était dans les années précédentes. De cette façon, il serait difficile de considérer le passage à une étude psychologique de la logique comme foncièrement moins psychologique qu'une étude psychologique de l'arithmétique, ce que l'auteur doit soutenir de quelque façon pour supporter la thèse selon laquelle le passage à l'étude des processus signitifs par Husserl dans la *Philosophie de l'arithmétique* de 1891 est un abandon de la psychologie descriptive et des prémisses psychologistes grevant une période brentanienne qu'il aurait commencée à dépasser en 1891, pour ainsi affirmer que l'apparition de la problématique signitive est le premier pas vers la phénoménologie en ce que celle-ci est foncièrement anti-psychologiste, ce que l'auteur n'affirme pas précisément mais qui est visiblement la conclusion de l'argumentation qu'il propose. D'autre part, il est étonnant de voir que Brisart tente de défendre une lecture anti-frégéenne de cette portion de l'œuvre de Husserl, en faisant abstraction des travaux suscités par cette perspective : nous pensons ici aux polémiques suscitées par Føllesdal, au sujet desquelles ont notamment pris parti Mohanty, Sokolowski et Willard, en défendant une lecture du premier Husserl à propos de laquelle il aurait été intéressant de voir Brisart prendre explicitement position. Dans la suite de Føllesdal, Smith et McIntyre ont insisté quant à eux sur

l'explication du concept de noème dans les mêmes termes que la notion fré géenne de sens (*Sinn*), et les différentes polémiques suscitées par les premiers travaux de Føllesdal auront contribué dans une très large mesure à réinvestir le champ des recherches husserliennes partant du thème de la signification, ce qui est tout simplement passé sous silence dans l'article de Brisart.

Dans son long article sur « Les trois étapes de la formation de la théorie de la signification chez le premier Husserl » (p. 63-218), English retrace l'évolution du thème de la signification de la *Philosophie de l'arithmétique* aux *Leçons de 1908* en passant par les *Recherches logiques*. Tout comme Brisart l'avait fait, English insiste également sur le fait que le manuscrit de 1890 est une première élaboration de la théorie de la signification que l'on retrouvera dans la *Philosophie de l'arithmétique* un an plus tard : partant de certaines prises de position psychologues sur la nature des nombres, Husserl était encore confronté, dans la *Philosophie de l'arithmétique*, à l'élaboration d'une théorie de la signification sur le modèle des concepts de nombres propres, et donc une théorie de la signification fonctionnant sur une modalité symbolique « au sens du remplacement ou de la substitution » (p. 65), modalité réductrice à celle de la perceptivité selon English, en ce que cette substitution ne permettait pas à l'intentionnalité de se dégager du type de structuration de la modalité perceptive de signification. Par cette modalité, English entend une définition de la signification en termes de désignation (*Bezeichnung*) à l'opposé de la signification à proprement parler (*Bedeutung*). Selon English, il y aurait aussi deux ontologies rattachées à cette théorie de la signification passant de la modalité perceptive à la modalité signitive, et donc de la première à la seconde étape annoncée par le titre de l'article : « Si la première théorie de la signification proposée par Husserl de 1887 à 1891, et exposée dans la *Philosophie de l'arithmétique*, avait donc souffert d'une grave insuffisance, c'était sans aucun doute à cause de cette absence de toute garantie ontologique qui aurait été aussitôt fournie par la référence à un pôle objectif véritablement transcendant » (p. 93). Aussi, le problème de cette première théorie de la signification réside selon English dans l'incapacité de Husserl à assurer un sol ontologique transcendant aux significations ou désignations de ce qui est donné dans la modalité perceptive. Mais English se fait vraisemblablement très dur, dans la mesure où la préoccupation de Husserl dans la *Philosophie de l'arithmétique* n'est pas tant de fournir une théorie de la signification telle qu'elle sera entendue dans l'optique réaliste des *Recherches logiques*, où même à la lumière des *Leçons de 1908*, mais plutôt de définir le statut des nombres et leur signification comme entités psychologiques puis, plus tard, comme signes. Mais qu'en est-il de cette seconde étape dans le développement de la théorie de la signification husserlienne? Selon English, Husserl serait parti dans la première étape de sa théorie de la signification par l'affinité originaire entre les vécus intérieurs et l'ontologie matérielle pour s'élever ensuite aux vécus supérieurs et à l'ontologie formelle, puis en redescendant vers le bas tout en conservant les résultats établis par le processus abstraitif (p. 95-96), en conservant donc de cette façon un parallélisme entre le matériel et le formel, mais en ne changeant pas de modalité malgré le changement de régime. Autrement dit, la signification demeurerait sur le niveau de la modalité perceptive, bien que les ontologies atteintes auraient exigé, selon English, une modalité différente, plus « objective ». Ce n'est toutefois pas ici la première étape telle qu'envisagée par English car, selon lui, c'est par des changements dans les modalités mêmes de l'intentionnalité qu'une théorie de la signification peut être amenée à se modifier. Il s'agit alors du mode d'exercice de l'intentionnalité, qui pour les concepts formels doit ajuster son tir, et nommément passer à un mode signitif.

Mais cette évolution de la première étape (*Philosophie de l'arithmétique*) à la seconde (*Recherches logiques*) marque également le développement d'un motif qui, bien que English n'emploie pas le mot, pourrait être qualifié de dialectique, en ce que le jeu qui s'institue entre les facteurs subjectifs qui relèvent du signifier et les facteurs objectifs que forment les significations crée « une espèce d'aller et retour perpétuel, dont l'étude devrait donc constituer désormais, et selon un régime de conditionnement bilatéral, s'exerçant sur une zone mixte, fondée des deux côtés, le centre même de toute théorie possible de la signification » (p. 125). En ce sens, il faut comprendre que « si la théorie de la signification doit donc trouver désormais son centre de gravité, ce ne peut être que sur la zone médiane même où doivent venir interférer ces deux séries dissymétriques de variations, les unes offertes par cet énorme pouvoir de modification dont l'intentionnalité se trouve investie subjectivement, en ayant toujours les moyens de changer ses genres de rapports aux objets qu'elle peut viser au-delà d'elle-même selon les différents actes par lesquels elle décide de les rejoindre, et les autres qui sont fournies, elles, par toutes les possibilités de combinaisons inhérentes aux structures essentiellement syncatégorématiques du langage avec les lois de transformation grammaticales qui leur sont propres » (p. 129). C'est ce qui expliquerait le troisième moment de la théorie husserlienne de la signification : il y aurait en 1908 un « renversement de perspective qui déplace le centre de gravité de toute cette problématique de l'ontologie vers la phénoménologie » (p. 143). Cependant, cette étape maintient en place les distinctions des deux premières étapes, mais va aussi les soumettre à une série de réélargissements caractéristiques, en procédant toutefois à une grande avancée, consistant à ne pas confondre les modes de visée signifiants et perceptifs. Enfin, ce problème de la théorie de la signification, bien qu'il semble achevé et substantiel dans ce texte de 1908, ne se distingue jamais du problème du développement de l'intentionnalité, trouble que manifeste le manuscrit de 1911 sur les significations empiriques, qui vient remettre en question la légitimité d'une systématique des actes intentionnels dans le domaine déictique.

Contrairement à Brisart, English se limite à un exposé plus descriptif qu'explicatif, bien que celui-ci vise à soutenir que le développement des thèmes de la phénoménologie transcendantale est effectif dès 1894, certes à travers la polémique entre Husserl et Twardowski sur le contenu des représentations, mais également en affirmant qu'il y avait dès cette époque un enjeu phénoménologique dans le passage d'une modalité perceptive à une modalité signitive (p. 193 et suivantes) pour la signification, ce qui peut être une thèse valable, mais qui mériterait pour cela d'être étudiée à la lumière d'une définition plus précise de la phénoménologie, ce que English croit d'emblée acquis chez ses lecteurs.

Enfin, le troisième et dernier article est celui de J. Benoist (« L'identité d'un sens : Husserl, des espèces à la grammaire », p. 219-271), qui tente de poursuivre une des idées développées dans son dernier livre (*Phénoménologie, sémantique, ontologie*), selon laquelle les *Recherches logiques* de Husserl se déploient dans un champ métaphysique neutre dépassant, d'une part, le débat idéalisme/réalisme, et ne manifestant, d'autre part, aucun engagement métaphysique. Cette idée se poursuit ici sur la base de la relation entre la grammaire et la signification, comme quoi la signification telle qu'elle est définie dans les *Recherches logiques* se complète par la grammaire, car comme Benoist le remarque, « le problème est celui de la parole elle-même comme contexte, de l'occasion inhérente à la parole qui se manifeste dans ce phénomène fondamental qui est celui de son indexicalité, phénomène que Husserl reconnaît

assurément comme fondamental, et qui l'embarasse fort, justement parce que lui se situe dans son analyse à ce niveau même qui est celui des actes du signifier, qui est aussi le niveau de cette occasionnalité » (p. 225).

Benoist insiste ici sur un problème auquel sera confronté Husserl de 1901 à 1908 quant à sa théorie de la signification, soit la place des significations empiriques. Si la signification est itérée dans l'acte, elle existe en un certain sens indépendamment de lui, au sens où l'acte n'est pas la signification, mais dans la mesure où la signification est obtenue dans l'acte, ce qui n'est pas sans rappeler un certain réalisme platonicien en ce que Husserl discute des significations comme d'entités, mais qui dans les faits n'a que très peu à faire avec le réalisme, en ce que les actes doivent être itérés, performés, et que cette clause performative a plus à faire avec un certain pragmatisme qu'avec un réalisme. Benoist ajoute : « Dans les *Recherches logiques*, le problème demeure de penser les formes d'unité d'une intentionnalité — celle de la signification — qui n'est pas tournée vers sa propre unité, mais vers celle de l'objet, qui est visé — même si celle-ci bien sûr ne pourra s'attester que dans celle-là, qui y constitue la voie d'accès obligée. La réponse tient dans la théorie des espèces de signification, qui ne sont rien d'autre alors que les formes d'unité de la signification » (p. 228).

L'auteur fait ici un parallèle entre le rouge *in specie* et les différents moments de rouge sur des bandes de papier rouge : si nous pouvons considérer la signification comme le rouge *in specie*, comme Husserl tente à première vue de nous le faire croire, on retrouverait la même signification dans différents énoncés comme le même rouge dans différents objets. Mais comme je ne vois pas le rouge comme je vois la signification, cette différence appuierait en un certain sens la distinction des genres signitifs, perceptifs et abstractifs dans la phénoménologie. Mais comme Benoist le mentionne en faisant référence à Rigal, le problème de l'identité sensible se rapproche de celui de l'identité de la signification, en ce que le caractère fondamentalement sémantique du problème de l'identité sensible s'y révèle comme n'étant pas celui de la perception de nuances du « même » rouge mais comme étant celui de la perception de nuances différentes de rouges comme même ment rouges : « Cette généralité [de la signification], en tant qu'idéalité, n'est généralité d'aucune généralisation (communisation), mais bien plutôt d'un manque, d'une absence d'inhérence à aucun acte qui est pourtant constitutive du sens de ces actes mêmes, et, faut-il dire rien que cela — c'est ce qui fait le statut irréductiblement phénoménologique de la signification dans les *Recherches logiques*, que critiquera l'exposé de 1908 » (p. 235). L'auteur conclut cette première partie en insistant sur le fait que les entités idéales que sont les significations dans les *Recherches logiques* ne sont pas des idéalités platoniciennes. Bien qu'il s'agisse d'entités idéales, elles sont d'abord et avant tout opératoires.

En deuxième partie, l'auteur met de l'avant la nécessité manifeste dans les *Leçons de 1908* de donner une certaine consistance ontologique à la signification : « C'est donc au moment où le platonisme de la signification au sens bolzanien du terme s'accuse que l'on est peut-être le plus loin du platonisme comme doctrine de l'idée. Il n'y a pas de transcendance où le sens pourrait déployer son manque, mais juste le plein, à vrai dire absolument vide (au sens de vide d'intuition, indifférent à elle) des formes de signification, dans leur conjugaison possible » (p. 258). Cette « objectivation » de la signification amène alors Husserl à radicaliser la distinction entre objets idéaux et objets empiriques, rendant la signification empirique inséparable du monde où elle apparaît, et ne la rabattant pas sur l'absence de limites de la raison objective, comme c'était le cas dans les *Recherches logiques*. Sans prendre

explicitement position sur le sens à donner ainsi que sur la valeur des différentes modifications que Husserl applique à sa théorie de la signification, il faut toutefois noter que l'article de Benoist prend des précautions louables quant à une explication de ces modifications « en zigzag », telle que la suggère trop allusivement English.

Enfin, il est dommage que le premier ouvrage en langue française sur la litigieuse question de la signification dans la première portion de l'œuvre de Husserl n'ait pas du tout tenté de comprendre le rôle joué par la psychologie descriptive dans le développement des *Recherches logiques* ainsi que dans son rapport à la théorie de la signification, une méthode que tire Husserl de l'enseignement de Brentano et qui motive sans doute plus qu'on ne le pense l'élaboration de sa phénoménologie dans la période couverte par l'ouvrage étudié ici.

GUILLAUME FRÉCHETTE
Université du Québec à Montréal

Livres reçus (printemps 2000)

Le matérialisme contemporain

Volume 27, numéro 1, printemps 2000

URI : <https://id.erudit.org/iderudit/004953ar>

DOI : <https://doi.org/10.7202/004953ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer ce document

(2000). Livres reçus (printemps 2000). *Philosophiques*, 27 (1), 227–228.
<https://doi.org/10.7202/004953ar>

Livres reçus (printemps 2000)

- Antonioni, Manola, 1999. *Deleuze et l'histoire de la philosophie*. Paris, Kimé, 122 p.
- Atlan, Henri, 1999. *Les étincelles du hasard*. Connaissance spermatique, tome 1, Paris, Seuil (coll. « La librairie du xx siècle »), 392 p.
- Bouton, Christophe, 2000. *Temps et Esprit dans la philosophie de Hegel. De Francfort à Iéna*, Paris, Vrin, (coll. « Temps modernes »), 319 p.
- Carrilho, Manuel Maria, 1999. *Rhétoriques de la modernité*. Paris, PUF, (coll. « Quadrige »), 175 p.
- Cléro, Jean-Pierre, 2000. *Théorie de la perception. De l'espace à l'émotion*. Paris, PUF, (coll. « L'interrogation philosophique »), 319 p.
- Delacampagne, Christian, 2000. *Histoire de la philosophie au xx^e siècle*. Paris, Seuil, (coll. « Points essais »), 404 p.
- Demonet, Marie-Luce (dir.), 1999. *Montaigne et la question de l'homme*. Paris, PUF, (coll. « Débats philosophiques »).
- Dortier, Jean-François (dir.), 1999. *Philosophies de notre temps*. Auxerres, Sciences humaines, 359 p.
- Dumouchel, Daniel, 1999. *Kant et la genèse de la subjectivité esthétique*. Paris, Vrin, (coll. « Histoire de la philosophie »), 305 p.
- Edelman, Bernard, 1999. *Nietzsche. Un continent perdu*. Paris, PUF, (coll. « Perspectives critiques »), 366 p.
- Gagnon, Maurice et Hébert, Daniel, 2000. *En quête de science. Introduction à l'épistémologie*. Saint-Laurent, Fides, 309 p.
- Gouin, Jean-Luc, 1999. *Hegel ou la Raison intégrale suivi de Aimer penser mourir : Hegel, Nietzsche, Freud en miroir*, Montréal, Bellarmin, 225 p.
- Grillo, Éric, 2000. *Intentionnalité et signifiante : une approche dialogique*. Bern, Peter Lang, 392 p.
- Grondin, Jean, 1999. *Introduction à Hans-Georg Gadamer*. Paris, Cerf (coll. « La nuit surveillée »), 232 p.
- Guilhaumou, Jacques, 1998. *La parole des sans. Les mouvements actuels à l'épreuve de la Révolution française*. Fontenay, éditions ENS, (coll. « Feuilles »), 140 p.
- Jaulin, Annick, 1999. *Aristote. La métaphysique*. Paris, PUF, (coll. « Philosophies »), 127 p.
- Koula Mellos et Patrick Savidan (dir.), 1999. *Pluralisme et délibération. Enjeux en philosophie politique contemporaine*. Ottawa, Presses de l'Université d'Ottawa, 193 p.
- Kistler, Max, 1999. *Causalité et loi de la nature*. Paris, Vrin, (coll. « Mathesis »), Paris, 311 p.
- Lazzeri, Christian, 1999. *Spinoza. Puissance et impuissance de la raison*. Paris, PUF, (coll. « Débats philosophiques »), 115 p.

- Lonergan, Bernard, 2000. *La compréhension de l'être*. Saint-Laurent, Bellarmin, 422 p.
- Machiavel, Nicolas, 2000. *Le Prince*. traduction et commentaire de Jean-Louis Fournel et Jean-Claude Zancarini, Paris, PUF, (coll. « Fondements de la philosophie politique »), 640 p.
- Mazzara, Giuseppe, 1999. *Gorgia/La Retorica del Verosimile*. Siegburg, Academia Verlag.
- Nadeau, Robert, 1999. *Vocabulaire technique et analytique de l'épistémologie*. Paris, PUF, (coll. « Premier cycle »), 872 p.
- Panaccio, Claude, 1999. *Le discours intérieur de Platon à Guillaume d'Occam*. Paris, Seuil, (coll. « Des travaux »), 341 p.
- Renaud, François, 1999. *Die Resokratisierung Platons*. Siegburg, Academia Verlag.
- Talbot, Gilbert, 1999. *Phil et Sophie ou de l'être humain*. Québec, Le loup de Gouttière, 84 p.
- Venet, Gisèle (dir.), 1998. *Le mal et ses masques. Théâtre, imaginaire, société*. Fontenay, éditions ENS, (coll. « Signes »), 458 p.
- Zancarini, Jean-Claude, 2000. *Le Droit de résistance, XII^e-XX^e siècles*. Fontenay, éditions ENS, (coll. « Theoria »).

Les personnes désireuses de faire un compte-rendu de l'un ou l'autre des ouvrages qui apparaissent dans cette liste sont priées de s'adresser à :

Daniel Dumouchel
Université de Montréal
Département de philosophie
C.P. 6128 succ. Centre-ville
Montréal (Québec), H3C 3J7
Tél.: (514) 343-6484
Télec.: (514) 343-7899
Courriel : dumouchd@netcourrier.com