

La théorie autoreprésentationnelle de la conscience¹

Uriah Kriegel

Trad. fr. D. Seron, dans A. Dewalque & Ch. Gauvry (éds.), *Conscience et représentation : Introduction aux théories représentationnelles de l'esprit*, Paris, Vrin, 2016, p. 201-224.

1. Une théorie de la conscience : premier survol

Cet ouvrage présente une théorie de la conscience². Je me livre dans cette toute première section à un premier survol de la théorie, dans la suite du chapitre à un deuxième survol, et dans le reste du livre à un troisième survol.

Quand j'ai une expérience consciente du ciel bleu, il y a un certain effet que ça me fait d'avoir cette expérience. Dans ce cas particulier, il y a un effet bleuique que ça me fait d'avoir cette expérience. Cet « effet effet bleuique que ça me fait » constitue le caractère phénoménal de mon expérience. Le caractère phénoménal est la propriété qui fait d'un état phénoménalement conscient (i) l'état phénoménalement conscient qu'il est et (ii) un état phénoménalement conscient en général.

On peut distinguer deux composantes de l'effet bleuique que ça fait pour moi : (i) la composante *bleuique* et (ii) la composante *pour-moi*. J'appelle la première « caractère qualitatif », la seconde « caractère subjectif ». En première approximation, le caractère phénoménal est juste la présence conjointe du caractère qualitatif et du caractère subjectif. En seconde approximation, il y a une division plus spéciale du travail conceptuel entre caractère qualitatif et caractère subjectif : le caractère qualitatif d'un état phénoménalement conscient est ce qui fait de lui l'état phénoménalement conscient qu'il est, tandis que son caractère subjectif est ce qui fait de lui un état phénoménalement conscient en général. Ainsi, mon expérience consciente du ciel bleu est l'expérience consciente qu'elle est en vertu de sa bleucité, mais elle est une expérience consciente en général du fait d'être pour moi.

Une théorie complète de la conscience phénoménale rendrait compte des deux caractères qualitatif et subjectif. Mais l'explication du caractère subjectif est plus importante

¹ « The Self-Representational Theory of Consciousness », ch. 1 de U. Kriegel, *Subjective Consciousness. A Self-Representational Theory*, Oxford, Oxford University Press, 2009, p. 1-21 (*N.d.T.*).

² Je traduis toujours « consciousness » par « conscience » et, faute de mieux, les termes « awareness » et « to be aware of » par « aperception » et « apercevoir » (*N.d.T.*).

philosophiquement. Car le caractère subjectif est ce qui fait que quelque chose est un état phénoménalement conscient *en général*, et le [2] mystère de la conscience est avant tout de comprendre pourquoi il y a *en général* de la conscience phénoménale.

La théorie de la conscience que je vais présenter tourne autour de l'idée que les états phénoménalement conscients ont un caractère qualitatif en vertu du fait qu'ils représentent des caractères environnementaux, et un caractère subjectif en vertu du fait qu'ils se représentent eux-mêmes. Plus spécialement, le caractère qualitatif d'un état conscient consiste à représenter certaines propriétés responsives [*response-dependent properties*] que possèdent les objets externes ; son caractère subjectif consiste à se représenter lui-même de manière appropriée. Aussi mon expérience du ciel bleu représente-t-elle à la fois le ciel et elle-même. Elle tire son caractère qualitatif (bleuicité) du fait qu'elle représente la propriété qu'a le ciel d'être disposé à obtenir une certaine réponse chez des répondants appropriés ; elle tire son caractère subjectif (son être-pour-moi) du fait qu'elle se représente elle-même de manière appropriée.

Il s'ensuit que ce qui fait de quelque chose un état phénoménalement conscient (en général) est l'autoreprésentation appropriée. Le raisonnement est celui-ci : le caractère subjectif d'un état phénoménalement conscient est ce qui fait de lui un état phénoménalement conscient (en général) ; un tel état a un caractère subjectif en vertu du fait qu'il se représente lui-même de façon appropriée ; donc, il est un état phénoménalement conscient (en général) en vertu du fait qu'il se représente lui-même de façon appropriée. Dès lors, l'existence de la conscience phénoménale en général s'explique par le fait que certains états mentaux se représentent eux-mêmes de la bonne manière. Voilà ce que j'appelle « théorie autoreprésentationnelle de la conscience », ou plus brièvement « autoreprésentationalisme ».

2. Le concept de conscience phénoménale

On a souvent l'impression que, sous le titre de « conscience phénoménale », des théories de la conscience différentes ont eu tendance à viser des phénomènes différents. Pour certains, cela suggère que beaucoup de controverses entre théories de la conscience sont purement verbales, reposant sur des sens différents des termes employés. D'après cette conception « pluraliste », un authentique progrès réclamerait, avant toutes choses, que la pluralité de phénomènes connus sous le nom de « conscience » soient soigneusement distingués et catalogués de façon systématique¹.

¹ Le terme « pluralisme » est employé par N. Block (« On a confusion about the function of consciousness », *Behavioral and Brain Sciences*, 18, 1995, p. 227-247) en référence à sa propre distinction entre quatre différents

[3] Je souhaite marquer mon accord sur l'idée que des théories de la conscience différentes ont visé des phénomènes différents, et pourtant insister sur le fait que l'attitude pluraliste est quelque peu fallacieuse. Ainsi, il est naturel de penser que ceux qu'on appelle les représentationalistes ont effectivement produit une théorie de quelque chose comme ce que j'ai appelé le caractère qualitatif, tandis que les partisans des théories d'ordre supérieur ont eu en vue quelque chose comme ce que j'ai appelé le caractère subjectif¹. En même temps, bien

phénomènes parfois ciblés par la théorie de la conscience, et qu'il appelle « conscience phénoménale », « conscience d'accès », « conscience de *monitoring* » et « conscience de soi ».

¹ Les théories représentationnelles affirment que le caractère phénoménal d'un état conscient est constitué par (un aspect de) son contenu représentationnel. Pour les théories représentationnelles, voir principalement F. Dretske, *Naturalizing the Mind*, Cambridge (Mass.), MIT Press, 1995 ; M. Tye, « Visual qualia and visual content », dans T. Crane (éd.), *The Contents of Experience*, Cambridge, Cambridge University Press, 1992 ; Id., *Ten Problems of Consciousness*, Cambridge (Mass.), MIT Press, 1995 ; Id., *Consciousness, Color, and Content*, Cambridge (Mass.), MIT Press, 2000 ; Id., « Visual qualia and visual content revisited », dans D.J. Chalmers (éd.), *Philosophy of Mind*, Oxford, Oxford University Press, 2002. Une version plus complexe est développée dans S. Shoemaker, « Phenomenal character », *Noûs*, 28, 1994, p. 21-38 ; Id., « Self-knowledge and "inner sense" ». Lecture III : The phenomenal character of experience », *Philosophy and Phenomenological Research*, 54, 1994, p. 291-314 ; Id., « Introspection and phenomenal character », dans D.J. Chalmers (éd.), *Philosophy of Mind*, *op. cit.* Toutes les versions modernes trouvent dans une certaine mesure leur inspiration dans ce qu'on appelle la transparence de l'expérience : les seules propriétés introspectivement accessibles des expériences conscientes sont leurs propriétés représentationnelles (voir G. Harman, « The intrinsic quality of experience », *Philosophical Perspectives*, 4, 1990, p. 31-52). Les théories d'ordre supérieur affirment que le caractère phénoménal d'un état conscient est constitué par sa présence dans le contenu représentationnel d'un état d'ordre supérieur numériquement distinct. La version la plus élaborée de la théorie d'ordre supérieur est sans doute celle de David Rosenthal (« Two concepts of consciousness », *Philosophical Studies*, 94, 1986) p. 329-359 ; « A theory of consciousness », *Technical Report 40*, ZIF Report Zentrum für Interdisziplinäre Forschung, Bielefeld, 1990 ; « Thinking that one thinks », dans M. Davies & G.W. Humphreys (éds.), *Consciousness : Psychological and Philosophical Essays*, Oxford, Blackwell, 1993 ; « Explaining consciousness », dans D.J. Chalmers (éd.), *Philosophy of Mind*, *op. cit.* ; *Consciousness and the Mind*, Oxford, Oxford University Press, 2005). Pour d'autres versions, voir principalement D. Armstrong, *A Materialist Theory of the Mind*, New-York, Humanities Press, 1968 ; W. Lycan, *Consciousness and Experience*, Cambridge (Mass.), MIT Press, 1996 ; Id., « The Superiority of HOP to HOT », dans R. Gennaro (éd.), *Higher-Order Theories of Consciousness*, Philadelphia-Amsterdam, John Benjamins, 2004 ; P. Carruthers, « Natural theories of consciousness », *European Journal of Philosophy*, 6, 1998, p. 203-222 ; Id., *Phenomenal Consciousness*, Cambridge, Cambridge University Press, 2000 ; R. Van Gulick, « Inward and upward — reflection, introspection, and self-awareness », *Philosophical Topics*, 28, 2001, p. 275-305 ; Id., « Higher-order global states (HOGS) : An alternative higher-order model of consciousness », dans R.J. Gennaro (éd.), *Higher-Order Theories of Consciousness*, *op. cit.* ; Id., « Mirror mirror — is that all ? », dans U. Kriegel & K.W. Williford (éds.), *Self-Representational Approaches to Consciousness*, Cambridge (Mass.), MIT Press, 2006. Toutes ces théories, ou la plupart d'entre elles, sont inspirées par ce qu'on appelle le principe de transitivité : les états conscients sont des états dont nous sommes conscients (ou que nous apercevons [*we are aware of*]) (voir D. Rosenthal, « Two concepts of consciousness », *art. cit.*). Fait intéressant, l'œuvre de certains critiques acharnés des théories représentationnelles et d'ordre supérieur révèle une éloge (implicite ou explicite) de ces théories pour avoir au moins ciblé le bon phénomène. On peut le voir, par exemple, dans la critique par Block des théories représentationnelles (« Inverted Earth », *Philosophical*

qu'à un certain niveau les philosophes aient peut-être poursuivi des phénomènes différents, à un autre niveau ils se préoccupent d'un seul et unique phénomène : une propriété, quelle qu'elle soit, génère en eux un sentiment de mystère quand ils considèrent la nature de la conscience et sa place dans le monde. L'expression « conscience phénoménale » est maintenant largement utilisée comme une étiquette pour désigner cette propriété.

En gardant cela à l'esprit, je propose le régime terminologique suivant. Le terme « conscience » est, bien entendu, un terme courant de la vie quotidienne qu'on peut comprendre intuitivement. Nous pouvons dès lors utiliser cette compréhension intuitive en fixant la référence de l'expression technique « conscience *phénoménale* » au moyen de la description définie rigidifiée suivante : « la propriété F , qui est telle que, dans le monde réel, F est responsable du mystère de la conscience ». Cette description peut être précisée de plusieurs manières. Supposons par exemple que nous lisions « responsable » en un sens causal, et que nous interprétions « mystère » au sens où les faits de conscience ne sont pas déductibles des faits physiques. Nous pourrions alors proposer la formulation plus précise suivante : « La conscience phénoménale est la propriété F qui est telle que, dans le monde réel, F produit causalement (mettons dans un sujet convenablement réflexif) le sentiment que les faits de [4] conscience ne peuvent être déduits de faits physiques. » D'autres reformulations plus précises s'attacheront à des compréhensions différentes de certaines composantes de la description définie rigidifiée d'où nous sommes partis. Mais toutes interpréteront la conscience phénoménale comme cette propriété, quelle qu'elle soit, qui est la source du mystère entourant la conscience.

Je soutiens que c'est là la manière la plus précise et profitable d'avoir les idées claires sur la notion de conscience phénoménale. C'est la plus précise parce qu'elle capture la raison pour laquelle la notion est devenue largement employée dans les cercles philosophiques, et c'est la plus profitable parce qu'elle nous permet d'extraire le noyau essentiel de nombreux débats sur la conscience qui semblent mêler obscurément ce qui est essentiel et ce qui est simplement verbal.

Par exemple, quand des théoriciens représentationalistes de la conscience débattent avec des partisans de la théorie d'ordre supérieur, on a le sentiment qu'à un certain niveau ils se préoccupent de propriétés différentes, mais aussi qu'à un autre niveau il existe entre eux un authentique désaccord. La caractérisation de la conscience phénoménale proposée ici confirme et explique ce sentiment. Bien que les explications des représentationalistes et des partisans de la théorie d'ordre supérieur puissent avoir pour objets des propriétés différentes,

Perspective, 4, 1990, p. 52-79 ; « Mental paint and mental latex », *Philosophical Issues*, 7, 1996, p. 19-50) et dans la critique par Levine des théories d'ordre supérieur (*Purple Haze : The Puzzle of Consciousness*, Oxford, Oxford University Press, 2001). Dans notre présente terminologie, Block semble estimer que le phénomène à cibler est le caractère qualitatif, alors que Levine croit que c'est le caractère subjectif.

en ceci que les premiers ont en vue le caractère qualitatif et les seconds le caractère subjectif, il y a aussi entre eux un désaccord essentiel — à savoir sur la question de savoir laquelle de ces propriétés est responsable du mystère de la conscience. Les représentationalistes semblent attachés à l'idée que le caractère qualitatif est la propriété qui, dans le monde réel, est la source de mystère. Les défenseurs de la théorie d'ordre supérieur pensent que cette propriété est le caractère subjectif.

La notion de conscience phénoménale est parfois introduite en référence à la locution « l'effet que ça fait ». Bien qu'elle renferme quelque chose d'évocateur, il est sans doute peu probable qu'elle soit d'une grande aide pour le non-initié, et elle n'éclaire pas la fonction de l'expression « conscience phénoménale » dans le discours concerné. Plus important, elle n'est d'aucune aide pour extraire le noyau essentiel de controverses partiellement confuses, puisque les parties opposées sont typiquement en désaccord sur la question de savoir quels sont les types d'états dans lesquels cela fait un certain effet de se trouver. Ainsi, les partisans des théories d'ordre supérieur maintiennent typiquement qu'il n'y a aucun effet que ça fait de se trouver dans un état mental pourvu d'un caractère qualitatif et non d'un caractère subjectif, alors que les représentationalistes pensent qu'il y en a un. À ce niveau de nuance, les intuitions sur la manière appropriée d'appliquer la terminologie de l'effet que ça fait semblent résider en aval de la théorisation philosophique. Ce que je soutiens, c'est que le mieux est d'introduire la notion de conscience phénoménale dans les termes de la description définie rigidifiée mentionnée plus haut, [5] ou plus généralement en termes de propriété responsable du sentiment de mystère entourant la conscience.

Il faut bien noter ceci : le terme « mystère », dans la description définie, ne doit pas être compris comme dénotant un mystère *prima facie* plutôt qu'un mystère *ultima facie*. En parlant du mystère de la conscience, je n'entends pas insinuer que la conscience serait un mystère *ultime* — c'est-à-dire un mystère qu'une démystification finale ne pourra jamais dissiper¹. Ce que j'entends suggérer, en revanche, c'est qu'elle est un *authentique* mystère — c'est-à-dire que notre sentiment de mystère devant la nature de la conscience est rationnel et approprié, et qu'il n'est pas dû à la simple confusion. L'existence de cet authentique mystère *prima facie* peut être reconnue autant par des dualistes que par des physicalistes. David Chalmers, un dualiste, l'expose en se servant de ce qu'il appelle le « problème difficile » de la conscience, par opposition aux « problèmes faciles »². Joseph Levine, un physicaliste, use de

¹ L'idée que la conscience est un mystère ultime a été défendue, notamment, par Colin McGinn (« Can we solve the mind-body problem ? », *Mind*, 98, 1989, p. 349-366 ; « Consciousness and space », *Journal of Consciousness Studies*, 2 1995, p. 220-30 ; *The Mysterious Flame*, Cambridge (Mass.), MIT Press, 1999 ; *Consciousness and its Objects*, Oxford, Oxford University Press, 2004).

² D.J. Chalmers, « Facing up to the problem of consciousness », *Journal of Consciousness Studies*, 2, 1995, p. 200-219 ; Id., « The puzzle of conscious experience », *Scientific American*, 1995, p. 62-68 ; Id., *The Conscious Mind*, Oxford, Oxford University Press, 1996.

l'expression « fossé explicatif » pour désigner un problème voisin¹. Je discuterai la nature de ce mystère, et les relations entre ces intitulés, au chapitre 8 du livre. Pour l'instant, je souhaite seulement souligner qu'abstraction faite de motivations plus larges, de nombreux philosophes ont le sentiment que la conscience renferme un certain mystère qu'on ne trouve pas dans d'autres aspects de la nature, et qu'ils emploient l'expression « conscience phénoménale » en référence à ce qui produit ce mystère.

Le fond de vérité dans l'attitude pluraliste est que, dans la théorie de la conscience plus que partout ailleurs, des édifices explicatifs différents sont souvent associés à des conceptions différentes du phénomène en attente d'explication. Mais il ne s'ensuit pas — et c'est une erreur de supposer — que toutes les conceptions du phénomène à expliquer sont également bonnes. Au contraire, certaines identifient mal la source du mystère de la conscience, alors que d'autres l'identifient correctement. Aussi la théorie de la conscience est-elle essentiellement un programme d'ensemble comprenant à la fois une analyse de l'*explanans* et une analyse de l'*explanandum*². Les deux sont censées *présenter les choses correctement*. L'analyse de l'*explanandum* est censée présenter les choses correctement du fait d'identifier correctement la source du mystère *prima facie* entourant la conscience ; l'analyse de l'*explanans* est censée [6] présenter correctement ce qui, ultimement, démystifie cette source. Cette manière de voir les choses, contrairement à l'approche pluraliste, implique que s'attarder sur l'*explanandum* n'est pas simplement une question de stipulation judicieuse, mais que cela engage des thèses essentielles. Corrélativement, il est rare que les débats sur la nature de la conscience soient purement verbaux, à supposer même qu'ils le soient jamais. Quand nous sommes tentés de dire qu'ils le sont, c'est parce que nous oublions que les controverses non verbales en théorie de la conscience peuvent avoir trait non seulement à la nature de l'*explanans*, mais aussi à la nature de l'*explanandum* – c'est-à-dire qu'elles peuvent porter sur la question de savoir quelle propriété est responsable du mystère de la conscience.

Si nous choisissons de caractériser la conscience phénoménale sur la base de la description définie rigidifiée mentionnée plus haut, le résultat sera probablement qu'il n'y a qu'une propriété dénotée par « conscience phénoménale ». Cette conception doit être distinguée (a) de la conception selon laquelle il y a *plus* d'une propriété de cette sorte, et (b) de la conception selon laquelle il y a *moins* d'une propriété de cette sorte (c'est-à-dire aucune).

¹ J. Levine, « Materialism and qualia : the explanatory gap », *Pacific Philosophical Quarterly*, 64, 1983, p. 354-361.

² Bien sûr, l'*explanandum* de toutes les théories de la conscience est par définition le caractère phénoménal, la propriété qui fait d'une expérience consciente (i) l'expérience consciente qu'elle est et (ii) une expérience consciente en général. Mais déjà au niveau suivant de la description, les analyses divergent sur la nature de cette propriété.

Pour qu'il y ait *moins* d'une propriété dénotée par « conscience phénoménale », il devrait n'y avoir aucune propriété systématiquement responsable de notre sentiment de mystère. Le sentiment de mystère devrait être provoqué en des occasions différentes par des propriétés sans rapports les unes avec les autres, ou bien être entièrement fondé sur un tissu de confusions. Bien que cohérente, cette position n'est *a priori* pas plausible : il conviendrait d'apporter un argument très fort si nous venions à la prendre au sérieux. Et en un sens, même ce scénario ne montrerait pas qu'il y a moins d'une propriété dénotée par « conscience phénoménale », mais seulement que la propriété, unique, est *disjonctive*.

Pour qu'il y ait *plus* d'une propriété dénotée par « conscience phénoménale », il devrait y avoir de multiples propriétés dont chacune est indépendamment responsable du sentiment de mystère qui s'abat sur nous quand nous considérons la manière dont la conscience s'intègre au monde naturel. Cela, une fois encore, n'est *a priori* pas plausible. Il est peut-être plausible que des propriétés multiples concourent à engendrer ce sentiment de mystère. Cependant cela ne signifierait pas que « conscience phénoménale » a plusieurs *denotata*, mais seulement qu'il a un *denotatum* conjonctif. Si, par exemple, le caractère qualitatif et le caractère subjectif doivent concourir à nous mystifier, alors le *denotatum* de « conscience phénoménale » est la propriété conjonctive d'être à la fois qualitatif et subjectif. Ce qu'il faudrait pour que « conscience phénoménale » ait plusieurs *denotata* (par exemple pour que l'expression soit ambiguë en référence au caractère qualitatif et au caractère subjectif), c'est que plusieurs propriétés distinctes suffisent [7] individuellement pour engendrer le mystère de façon surdéterminée (par exemple que les caractères qualitatif et subjectif nous mystifient séparément et indépendamment)¹. Ce n'est pas absolument impossible, mais ce n'est pas non plus plausible *prima facie*. Et c'est seulement à ces conditions que l'attitude pluraliste décrite au début serait fondée.

Le point principal est qu'il est meilleur de concevoir les affirmations sur ce qu'il faut expliquer en théorie de la conscience comme des thèses essentielles sur ce qui engendre le mystère de la conscience, plutôt que comme de simples stipulations. En les prenant comme des stipulations, on ne fait qu'esquiver la source de l'angoisse philosophique entourant la conscience. On peut multiplier à l'infini les explications de notions stipulées différentes de la conscience, mais les seules à être dotées de sens pour cette angoisse philosophique seraient celles qui ciblent ce qui engendre en premier lieu le mystère de la conscience.

¹ Même dans ces conditions, on pourrait arguer que la conscience phénoménale devrait être identifiée à une unique propriété – à savoir à la propriété disjonctive qui est telle que chacun des termes de la disjonction soit une de ces sources surdéterminantes de mystère. Cette manière de voir, peut-on penser, est prescrite par l'emploi de l'article défini dans la description rigidifiée.

3. La structure du caractère phénoménal

En nous tournant maintenant vers la nature de l'*explanandum*, considérons mon expérience consciente alors que je regarde le ciel bleu par la fenêtre. Mon expérience a de nombreuses propriétés. La grande majorité d'entre elles sont complètement inintéressantes : elle se produit un mardi, 648 ans après la mort de l'homme le plus grand du quatorzième siècle, etc. Mon expérience possède ces propriétés, mais nous ne nous y intéressons pas, en partie parce que nous les comprenons déjà parfaitement bien.

Certaines des propriétés de mon expérience sont plus intéressantes. Ainsi, l'expérience a un fort impact sur la mémoire de travail ; elle induit une activation neuronale en V4 ; elle transmet de l'information sur le ciel¹. Ces propriétés largement psychologiques sont plus intéressantes, en partie parce que pour l'instant nous les comprenons imparfaitement. En même temps, nous avons une bonne idée de ce que devrait — ou devra — être une explication d'ensemble de ces propriétés. Nous ne sommes pas mystifiés par elles.

[8] L'existence d'un effet que ça me fait d'avoir mon expérience est plus proche de quelque chose qui peut véritablement nous mystifier². En l'occurrence, il y a un effet bleuique que ça me fait d'avoir cette expérience. Comme je l'ai dit au début, nous pouvons distinguer deux *aspects* ou *composantes* de cet « effet bleuique que ça me fait ». Il y a d'une part la composante *bleuique*, que j'appelle le caractère qualitatif de l'expérience, et d'autre part la composante *pour-moi*, que j'appelle le caractère subjectif de l'expérience³.

Quand un daltonien regarde le même ciel dans les mêmes conditions, il y a une différence psychologiquement réelle entre son expérience et la mienne. La dimension psychologiquement réelle sur laquelle nos expériences diffèrent est ce que j'appelle le caractère qualitatif. Nous n'avons pas besoin, à ce stade, de nous engager en faveur d'une

¹ Bien entendu, certains philosophes peuvent nier n'importe laquelle de ces affirmations, mais il est prématuré, à ce stade, de discuter de doctrines philosophiques particulières du genre impliqué ici. J'énumère ces propriétés à titre de propriétés raisonnables qu'on peut s'attendre à trouver dans les états conscients, au moins dans une attitude pré-philosophique.

² La locution, dans son usage actuel, est due à B.A. Farrell (« Experience », *Mind*, 59, 1950, p. 170–198), mais on doit à T. Nagel (« What is it like to be a bat ? », *Philosophical Review*, 83, 1974, p. 435–450) de l'avoir largement popularisée. Comme je l'ai dit dans la section précédente, l'expression n'est peut-être pas d'une grande utilité pour introduire la notion de conscience phénoménale auprès du non-initié. Mais je m'adresse ici à l'initié, et de toute façon je ne suis pas (encore) en train d'affirmer que l'effet que ça me fait est la même chose que la conscience phénoménale.

³ On peut se demander si le qualitatif et le subjectif sont réellement séparables. Même s'ils ne sont pas séparables dans la réalité, cependant, ils sont certainement séparables « en pensée ». C'est-à-dire qu'il y a ici à tracer une distinction conceptuelle même s'il n'y correspond aucune distinction entre propriétés. J'en dirai davantage sur ce point au chapitre 2 du présent ouvrage.

conception déterminée du caractère qualitatif. Sans prétendre fournir par là une explication, disons seulement ceci : les propriétés qualitatives « correspondent » aux caractères sensibles de l'environnement¹. Pour toute propriété dans l'environnement d'une créature donnée dont celle-ci peut avoir une expérience sensible, il y a une propriété qualitative potentielle d'une expérience consciente de cette créature. Le caractère qualitatif d'une expérience à un instant donné peut être conçu comme l'ensemble ou la somme de toutes les propriétés qualitatives que l'expérience instancie à cet instant.

En ce qui concerne le caractère subjectif, dire que mon expérience possède un caractère subjectif revient à pointer du doigt une certaine *aperception* [*awareness*] que j'ai de mon expérience. Les expériences conscientes ne sont pas des états qui peuvent, pour ainsi dire, passer inaperçus quand nous les *abritons*. Les états refoulés de Freud, les états subpersonnels et quantité d'autres états inconscients peuvent survenir en nous sans que nous en ayons le moins du monde connaissance, mais notre intuition est que les expériences conscientes sont différentes. Un état mental complètement inaperçu n'est pas une expérience consciente. En ce sens, mon expérience consciente n'est pas seulement *en moi*, elle est aussi *pour moi*².

[9] Une raison pour laquelle il est utile de distinguer entre caractères qualitatif et subjectif est que, comme on l'a déjà noté, les théories philosophiques de la conscience ont eu tendance à se répartir en deux groupes, l'un ciblant le caractère qualitatif, l'autre le caractère subjectif³.

Plusieurs conceptions sont possibles au sujet des interrelations entre caractères phénoménal, qualitatif et subjectif. Ce point sera discuté en détail au chapitre 2 du présent ouvrage. Pour l'instant, il suffit de dire que quatre positions se détachent plus centralement. L'une est de dire que le caractère phénoménal est identique au caractère qualitatif ; ce semble être la conception, par exemple, des représentationalistes typiques. D'après une autre conception, le caractère phénoménal est identique au caractère subjectif ; ce semble être la conception, par exemple, des partisans des théories d'ordre supérieur. Une autre conception encore consiste à dire que le caractère phénoménal est identique à quelque chose comme la comprérence ou la conjonction des caractères qualitatif et subjectif. Enfin, il y a aussi la conception suivant

¹ Nous pouvons ici interpréter l'environnement de manière à y inclure le corps du sujet.

² Il y aurait beaucoup plus à dire de la nature du caractère subjectif, ou sur ce point du caractère qualitatif. Et on en dira davantage au chapitre 2 du présent ouvrage. D'autre part, je trouve que quoi qu'on dise dans ce domaine, c'est toujours contestable et cela se ramène, dans cette mesure, à une thèse essentielle. Pour cette raison, je ne souhaite pas m'étendre sur la nature des caractères subjectif et qualitatif dans les présents développements, qui visent seulement à introduire les deux notions. J'en dis juste assez pour donner au lecteur une idée de ce dont je parle.

³ Ainsi, les théories représentationnelles semblent cibler le caractère qualitatif, alors que les théories d'ordre supérieur semblent cibler le caractère subjectif. Comme je l'ai dit dans la section précédente, on ne doit pas voir dans ce gouffre entre les deux approches une simple controverse verbale, mais une controverse essentielle portant sur la source du mystère de la conscience.

laquelle la conscience phénoménale n'entretient pas de relations constitutives avec le caractère qualitatif ou subjectif.

Deux conceptions me paraissent séduisantes. La première est moins plausible mais elle est très claire : le caractère phénoménal est identique au caractère subjectif (c'est-à-dire que le caractère subjectif est la source du mystère de la conscience). La seconde est plus plausible mais aussi plus compliquée : le caractère phénoménal est une sorte de propriété complexe incluant et le caractère subjectif et le caractère qualitatif. Il est difficile d'argumenter sur ces sujets, bien que je projette de les discuter plus en détail au chapitre 2 et dans l'Appendice. Pour le moment, et simplement pour marquer les différences de sensibilité qui se cachent derrière les conceptions qui me paraissent séduisantes, nous pouvons considérer ceci.

S'agissant de la première conception, on peut soutenir que l'(authentique) perplexité suscitée par le fait qu'une grappe de neurones vibrant à l'intérieur du crâne s'associe à un caractère qualitatif jaunique n'est pas différente de la perplexité entourant le fait qu'une grappe d'atomes allant et venant dans le vide s'associe à une couleur jaune. Que des neurones puissent être au fondement de la jaunicité, cela n'est pas plus étonnant que le fait que des atomes puissent être au fondement de la couleur jaune¹. Et pourtant le second n'est pas profondément [10] mystérieux. Par contre, le mystère entourant le fait qu'une grappe de neurones vibrant à l'intérieur du crâne s'associe à un être-pour-moi [*for-me-ness*] n'a pas de parallèle dans le domaine non mental. Il est surprenant que des neurones puissent être au fondement de l'être-pour-moi, et il n'y a pas de surprise analogue du côté de ce au fondement de quoi peuvent être les atomes. Aussi ce phénomène est-il plus adapté si l'on veut capturer le mystère distinctif que présente la conscience².

Quant à la deuxième conception, elle m'attire si je considère que, quand nous disons qu'un état mental est phénoménalement conscient, il doit y avoir (a) quelque chose qui fait de lui l'état phénoménalement conscient *qu'il est*, et (b) quelque chose qui fait de lui un état phénoménalement conscient *en général*. Pour la raison invoquée au paragraphe précédent et pour d'autres raisons encore, je pense que le caractère subjectif est ce qui rend un état mental phénoménalement conscient *en général*. Mais cela laisse quelque peu dans l'obscurité ce qui

¹ Cette observation est due à S. Shoemaker (« Self-knowledge and "inner sense". Lecture II : The broad perceptual model », *Philosophy and Phenomenological Research*, 5, 1994, p. 271-290). Il faut noter, cependant, que certains philosophes ont fait la même observation en vue de soutenir la thèse inverse : le fossé explicatif entre conscience et matière neuronale n'est rien d'autre que le fossé explicatif entre l'image manifeste des objets externes et leur image scientifique (A. Byrne, « Color and the Mind-Body Problem », *Dialectica*, 60, 2006, p. 223-244).

² Cette remarque est hâtive et peu soignée, et je n'entends pas m'y apesantir à ce stade. C'est simplement une manière d'exprimer ce à quoi je tends dans ce domaine. Une approche plus systématique des difficultés apparaissant ici sera mise en avant à la fin du chapitre 2 du présent ouvrage.

fait de lui l'état phénoménalement conscient *qu'il est*, une fois admis qu'il est phénoménalement conscient en général. Assez naturellement, on pensera que c'est le caractère qualitatif de l'état qui fait de lui l'état phénoménalement conscient qu'il est. Le résultat est l'idée que le caractère qualitatif fournit les conditions d'identité de la phénoménalité, tandis que le caractère subjectif en fournit les conditions d'existence : ce qui fait d'un état mental un état phénoménalement conscient en général (plutôt qu'un état non phénoménal), c'est son caractère subjectif ; ce qui fait de lui l'état phénoménalement conscient qu'il est (plutôt qu'un autre), c'est son caractère qualitatif.

D'après cette thèse sur la division du travail conceptuel, le caractère qualitatif de mon expérience (sa bleucité) détermine *quel* état phénoménalement conscient il est, cependant que son caractère subjectif (son être-pour-moi) détermine *qu'il est* un état phénoménalement conscient. Ainsi, pour qu'un état mental *devienne* phénoménalement conscient, il lui faut acquérir un caractère subjectif ; une fois qu'il *est*, pour ainsi dire, devenu phénoménalement conscient, c'est son caractère qualitatif qui fait de lui le type spécifique d'état phénoménalement conscient qu'il est¹.

[11] Comme je l'ai indiqué, une particularité assez curieuse de cette position est qu'elle dissocie complètement les conditions d'identité et les conditions d'existence. Typiquement, ce qui fait de quelque chose le *F* qu'il est et ce qui fait de lui un *F* en général ne sont pas complètement sans relation l'un avec l'autre, mais ils sont plutôt reliés comme un déterminé au déterminable. Ce qui fait de quelque chose une voiture en général est qu'il est une machine automobile (avec en plus quelques fioritures), tandis que ce qui fait de lui la voiture qu'il est est le fait qu'il est une machine automobile *de telle ou telle marque*. La propriété d'être une machine automobile de telle ou telle marque est le déterminé de la propriété d'être une machine automobile. De même, si le caractère qualitatif bleuique de mon expérience est ce qui fait d'elle l'état phénoménalement conscient qu'elle est, alors nous devrions nous attendre à ce que la possession de *tel ou tel* caractère qualitatif fasse d'elle un état phénoménalement conscient en général. Ici, le caractère subjectif ne joue absolument plus aucun rôle — ce qui n'est guère plausible.

La solution est de concevoir le caractère clé de mon expérience consciente comme étant l'*être-bleuique-pour-moi* [*bluish-for-me-ness*] (plutôt qu'un caractère factorisable en deux composantes, bleucité et être-pour-moi). D'une certaine manière, cela revient à soutenir que le caractère subjectif et le caractère qualitatif ne sont pas des propriétés séparées, mais en un

¹ J'emploie ici le langage du processus de manière purement métaphorique. Il n'y a pas réellement de processus au cours duquel un état mental deviendrait d'abord une expérience consciente en général et ensuite — avec un décalage temporel — deviendrait un type particulier d'expérience consciente. Naturellement, les deux surviennent simultanément. En fait, à suivre les acceptions les plus raisonnables du mot « occurrence », les deux occurrences n'en sont qu'une.

certain sens (problématique, je l'admets) des aspects d'une unique propriété¹. En un sens assez séduisant, l'être-bleuque-pour-moi est un déterminé de l'être-pour-moi (ou de l'« être-quelque-chose-pour-moi », si vous préférez). Ce qui produit l'effet souhaité : les conditions d'existence et d'identité de la phénoménalité sont reliées comme l'est un déterminable à un déterminé.

Telle est la position que je favorise sur la structure de la conscience phénoménale, et je la développerai davantage au chapitre 2 de ce livre. Elle voit dans le caractère phénoménal une affaire de complexe comprérence du caractère qualitatif et du caractère subjectif, lesquels ne sont pas entièrement séparables. Partant, une théorie de la conscience phénoménale doit s'intéresser aux deux. Le chapitre 3 développe une analyse du caractère qualitatif et le chapitre 4 du caractère subjectif. Les chapitres suivants se focalisent sur des aspects du caractère subjectif. La raison en est que le modèle finalement obtenu réserve encore une place plus centrale au caractère subjectif, puisqu'il constitue les conditions d'existence de la phénoménalité. C'est important parce que, bien que le profond mystère de la conscience doive avoir un rapport avec le pourquoi et le comment des différences entre épisodes conscients, il s'attache bien davantage à la question de savoir pourquoi et comment il y a *d'emblée* des épisodes conscients — c'est-à-dire pourquoi il existe quelque chose de tel que la conscience *en général*².

[12] 4. Une explication représentationnelle étroite du caractère qualitatif

Le but d'une théorie réductive de la conscience phénoménale est d'expliquer le caractère phénoménal en termes non phénoménaux. Si les termes non phénoménaux employés dans l'explication sont purement physiques, le résultat est une théorie *physicaliste* de la conscience. Mais toute théorie réductive n'est pas par définition physicaliste. Ce qui définit une théorie réductive est qu'elle explique le phénoménal en termes non phénoménaux.

¹ Après tout, même si mon expérience du ciel est *pour moi*, tout aspect de mon expérience n'est pas pour autant pour moi au sens pertinent. Plausiblement, le seul aspect qui est pour moi au sens pertinent est sa bleuicité.

² Ce qui fait mystère, c'est avant tout le fait qu'il y a certains événements dans ce monde qui semblent catégoriquement différents des événements brutaux, aveugles, inanimés de la sphère inconsciente. Ainsi le mystère concerne d'abord ce qui fait de quelque chose un état conscient *en premier lieu*, et seulement de manière dérivée ce qui fait de lui l'état conscient qu'il est étant donné qu'il en est un. Cela suggère que le caractère subjectif doit être la cible centrale de la théorie de la conscience, dans la mesure où la démystification de la conscience réclamerait plus centralement la démystification du caractère subjectif.

Le but premier de cet ouvrage est de fournir une explication réductive et, espérons-le, physicaliste du caractère phénoménal. Une explication réductive du caractère phénoménal aurait deux parties : une explication réductive de ce qui fait d'un état phénoménalement conscient l'état phénoménalement conscient qu'il est, et une explication de ce qui fait d'un état phénoménalement conscient un état phénoménalement conscient en général. Compte tenu de la thèse sur la division du travail conceptuel qu'on a avancée dans la section précédente, la première partie serait constituée d'une explication réductive du caractère qualitatif, et la seconde d'une explication réductive du caractère subjectif.

Au chapitre 3, je promeus une explication représentationnaliste du caractère qualitatif : le caractère qualitatif d'une expérience consciente est constitué par (un aspect de) son contenu représentationnel. Selon une telle conception, quelles que soient les autres choses qu'elle représente par ailleurs, mon expérience du ciel bleu représente aussi des caractères du ciel. Naturellement, le ciel a beaucoup de caractères différents, et dès lors mon expérience instancie beaucoup de propriétés représentationnelles différentes en relation au ciel bleu. D'après l'analyse que je défends, seul un sous-ensemble très spécifique de ces propriétés représentationnelles est constitutif de son caractère qualitatif. En particulier, seule la représentation de certaines propriétés responsives [*response-dependent properties*] du ciel peuvent être constitutives du caractère qualitatif de mon expérience.

Comment allons-nous comprendre les propriétés responsives pertinentes ? Répondre à la question est une manière d'étoffer sa version particulière de *représentationalisme responsif* (c'est ainsi que j'appellerai cette conception). [13] Une grosse part du chapitre 3 se concentrera sur les propriétés responsives pertinentes. En tout cas, un résultat de cette variété de représentationnalisme responsif est qu'il voit dans le contenu représentationnel constitutif du caractère qualitatif un *contenu étroit*¹. Les expériences conscientes peuvent bien avoir des propriétés représentationnelles larges, mais celles-ci ne sont pas constitutives de leur caractère qualitatif.

Ma principale préoccupation dans cet ouvrage, cependant, portera sur l'explication réductive du caractère subjectif. Il y a deux séries de raisons à cela. Les raisons inessentiels sont (a)

¹ J'ai recours ici à la distinction, due originellement à H. Putnam (« The meaning of "meaning" », dans *Mind, Language, and Reality*, Cambridge, Cambridge University Press, 1975), entre contenu étroit et contenu large. Le contenu étroit est le contenu qui est complètement déterminé par les propriétés non relationnelles du sujet de l'état dont il est le contenu. Le contenu large est le contenu qui n'est pas déterminé de cette manière. C'est du moins une manière parmi d'autres de présenter cette distinction. Le représentationnalisme étroit au sujet du caractère qualitatif, dès lors, est la conception suivant laquelle le caractère qualitatif d'une expérience consciente est constitué par des aspects du contenu de l'expérience qui sont complètement déterminés par les propriétés non relationnelles du sujet. J'ai pour la première fois défendu une version de cette conception (sous le titre de « représentationnalisme internaliste ») dans « Phenomenal content », *Erkenntnis*, 57, 2002, p. 175-198.

que la littérature m'apparaît de manière frappante beaucoup plus avancée quand il s'agit du caractère qualitatif que quand il s'agit du caractère subjectif, (b) qu'une conception représentationaliste très semblable à celle que je défendrai a déjà été défendue dans la littérature¹, et (c) que les idées nouvelles que j'ai à offrir portent pour la plupart sur le caractère subjectif. Les raisons essentielles ont à voir avec le fait que, d'après la conception défendue ici, le caractère subjectif est ce qui fait d'un état phénoménalement conscient un tel état en général, et que c'est là le facteur le plus profond du mystère philosophique entourant la conscience. C'est-à-dire que la condition d'existence de la phénoménalité (par opposition à sa condition d'identité) est l'aspect de la conscience qui réclame une explication de la manière la plus pressante, et le caractère subjectif est ce qu'on a besoin de comprendre si l'on veut fournir une telle explication.

5. Une explication autoreprésentationnelle du caractère subjectif

La thèse centrale de ce livre est que ce qui fait de quelque chose un état phénoménalement conscient en général, ce qui constitue son caractère subjectif, est une certaine espèce d'autoreprésentation : un état mental a un caractère phénoménal en général quand, et seulement quand, il se représente lui-même de la bonne manière. Tous les états phénoménalement conscients, et eux seuls, sont convenablement autoreprésentants. Ainsi, quelles que soient les autres choses qu'un état conscient représente par ailleurs, il se représente toujours lui-même, [14] et c'est en vertu du fait de se représenter lui-même qu'il est un état conscient. Voilà pour l'autoreprésentationalisme.

Bien qu'il ne soit pas sous les feux de la rampe dans les discussions modernes sur la conscience, l'autoreprésentationalisme a derrière lui une histoire vénérable. Un de ceux à l'avoir très tôt soutenu clairement et explicitement est Franz Brentano, qui écrit ceci dans sa *Psychologie d'un point de vue empirique* :

[Tout acte conscient] inclut en lui une conscience de lui-même. Pour cette raison, tout acte psychique [conscient], aussi simple soit-il, a un double objet, un objet primaire et un objet secondaire. L'acte le plus simple, par exemple l'acte d'entendre, a pour objet primaire le son, mais pour objet secondaire lui-même, le phénomène psychique dans lequel le son est entendu².

¹ Par exemple par S. Shoemaker, « Phenomenal character », art. cit.

² F. Brentano, *Psychologie vom empirischen Standpunkt*, Hamburg, Meiner, 1973, p. 218, trad. fr. par M. de Gandillac, revue par J.-F. Courtine, *Psychologie du point de vue empirique*, Paris, Vrin, 2007, p. 166. Brentano introduit pour la première fois cette idée dans les sections 7 et 8 du chapitre 2 (« Sur la conscience interne ») du livre 2. La section 8 porte le titre « Une représentation et une représentation de la représentation sont données dans un même et unique acte ». Dans cette section, Brentano expose sa conception des expériences conscientes comme autoreprésentationnelles.

Ainsi, l'expérience auditive consciente du son d'une cornemuse est dirigée intentionnellement à la fois vers la cornemuse et vers elle-même. Sur le fond d'une conception représentationnelle de l'intentionnalité, cela nous obligerait à soutenir la thèse suivant laquelle tous les états conscients sont autoreprésentants¹. Il se peut que Brentano ait emprunté cette idée à Aristote. Bien qu'il soit loin d'être aussi explicite que Brentano, Aristote écrit en effet dans la *Métaphysique* que « le connaître, le croire et le penser [conscients] sont toujours un connaître, un croire et un penser de quelque chose d'autre, mais d'eux-mêmes sur le côté [*on the side*] »².

Mais qu'est-ce que cela veut dire, pour un état mental, se représenter soi-même ? Répondre à cette question est une manière d'étoffer l'autoreprésentationalisme. Des versions différentes de cette conception produiront des réponses différentes. Je donnerai ma propre interprétation aux chapitres 4 et 6. En tout cas, jusqu'à preuve [15] du contraire, il n'y a aucune raison de suspecter qu'il y ait quelque chose de profondément inintelligible dans la notion d'état autoreprésentant.

Certaines expressions linguistiques révèlent des formes familières d'autoreprésentation³. L'exemple de phrase « cette phrase-ci est écrite en Garamond » est autoreprésentante. Le cas échéant, elle est une phrase vraie, et elle est elle-même une composante de son vérifacteur⁴. Si l'on suppose, de façon plausible, que pour être vraie une phrase doit être en contact sémantique avec son vérifacteur, et que le contact sémantique est réalisé — et peut-être même constitué — par la représentation des composantes du vérifacteur, alors cette phrase vraie doit

¹ On peut considérer une conception représentationnelle de l'intentionnalité soit comme une thèse tautologique, soit comme une thèse essentielle mais vraie, selon qu'on comprend ces termes dans un sens ou dans un autre. J'ai observé les deux dans la littérature. Dans la philosophie de l'esprit récente, il est plus courant de traiter une telle conception comme étant une tautologie, mais il y a des exceptions (par exemple R. Cummins, *Mental Representation*, Cambridge (Mass.), MIT Press, 1989).

² *Métaphysique*, Λ, 9, 1074b 35-36. J'utilise ici la traduction de V. Caston (« Aristote on consciousness », *Mind*, 111, 2002, p. 751-815). Un spécialiste au moins m'a suggéré que « sur le côté » [*on the side*] n'était pas la meilleure traduction, et que « secondairement » [*as a byproduct*] serait plus approprié. Il n'est pas clair que cela fasse une vraie différence pour notre propos. Il y a aussi des passages du *De Anima* qui peuvent être lus comme autoreprésentationalistes. Pour une interprétation autoreprésentationaliste approfondie d'Aristote, voir V. Caston, *ibid.* Caston note qu'on trouve aussi une telle interprétation d'Aristote dans une dissertation sur l'unité de la vie mentale dans la philosophie aristotélicienne, écrite sous la supervision de Brentano par un certain J. Herman Schell.

³ Il existe aussi des modèles mathématiques plus rigoureux de l'autoreprésentation, en particulier dans la théorie des ensembles non bien fondés (voir K.W. Williford, « The Self-representational structure of consciousness », dans U. Kriegel & K.W. Williford (éds.), *Self-Representational Approaches to Consciousness*, *op. cit.* ; G. Landini, « Consciousness and non-wellfoundedness », inédit).

⁴ Si l'on préfère concevoir les vérifacteurs comme des états de choses, on pourrait dire que le vérifacteur de cette phrase est l'état de choses consistant dans son instanciation de la propriété d'être écrit en Times New Roman.

représenter les composantes de son vérificateur, y compris elle-même^{1 2}. Je ne cite pas cette forme d'autoréférence de phrase comme un modèle fidèle pour décrire comment une autoreprésentation est rendue consciente. Je l'invoque simplement en manière de réponse à qui s'inquiéterait du fait que la notion d'autoreprésentation serait, pour une raison ou une autre, profondément inintelligible.

L'autoreprésentationalisme peut utilement être vu comme résultant de deux thèses. La première est que tous les états phénoménalement conscients sont conscients en vertu du fait d'être *représentés* (de la bonne manière). La seconde est qu'aucun état phénoménalement conscient n'est conscient en vertu du fait d'être représenté par un état *numériquement distinct* — c'est-à-dire par un état autre que lui-même. Il suit de ces deux thèses que tous les états phénoménalement conscients sont phénoménalement conscients en vertu du fait d'être représentés (de la bonne manière) par un état mental qui n'est pas numériquement distinct d'eux — c'est-à-dire en vertu du fait de se représenter eux-mêmes (de la bonne manière). En termes plus formels, nous pourrions énoncer l'argument principal comme suit. Pour tout état phénoménalement conscient *C*,

1. *C* est conscient en vertu du fait d'être représenté de manière appropriée ;
2. ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté par un état numériquement distinct ; donc, [16]
3. *C* est conscient en vertu du fait d'être représenté de manière appropriée par lui-même ; c'est-à-dire que
4. *C* est conscient en vertu du fait de se représenter lui-même de manière appropriée³.

L'argument est clairement valide, en sorte que les deux prémisses impliquent vraiment l'autoreprésentationalisme. La question est de savoir si elles sont vraies. Dans les deux sections suivantes j'esquisse pour chacune d'elles les principales raisons de la défendre.

¹ J'emploie l'expression « lui-même » comme un réflexif indirect. J'en dirai beaucoup plus au sujet des particularités sémantiques des réflexifs indirects au chapitre 4. Pour un travail précurseur sur ce sujet, voir H.-N. Castañeda, « "He" : A study in the logic of self-consciousness », *Ratio*, 8, 1966, p. 130-157 ; Id., « On the phenomeno-logic of the I », *Proceedings of the 14th International Congress of Philosophy III*, 1969, réimpr. dans Q. Cassam (éd.), *Self-Knowledge*, Oxford, Oxford University Press, 1994.

² J'insiste sur ce point.

³ Si je laisse ici et dans la suite le quantificateur en dehors de la présentation indentée de l'argument, c'est pour des motifs stylistiques. Mais il devrait être clair que la présentation dans le texte est équivalente à celle-ci :

1. Pour tout état conscient *C*, *C* est conscient en vertu du fait d'être représenté de manière appropriée ;
2. pour tout état conscient *C*, ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté par un état numériquement distinct ; donc
3. pour tout état conscient *C*, *C* est conscient en vertu du fait d'être représenté de manière appropriée par lui-même ; c'est-à-dire que
4. pour tout état conscient *C*, *C* est conscient en vertu du fait de se représenter lui-même de manière appropriée.

6. Conscience phénoménale et aperception interne

La première prémisse de l'argument principal de l'autoreprésentationalisme est la thèse qu'un état conscient est conscient en vertu du fait d'être représenté de manière appropriée. En termes impressionnistes, le motif général pour défendre cette prémisse est l'idée que, d'une manière ou d'une autre, il est essentiel à un état conscient que son sujet l'aperçoive. Les états conscients ne sont pas des états qui ne font qu'avoir lieu *en nous*, que nous nous en apercevions ou non ; ils sont aussi *pour nous*, précisément au sens où il y a un effet que ça fait d'être dans ces états. Des états mentaux qui surviennent simplement *en nous* mais dont nous n'avons aucune aperception ne sont pas des expériences conscientes¹.

Appelons *aperception externe* l'aperception de caractères et d'objets externes situés dans l'environnement ou le corps du sujet, et *aperception interne* l'aperception d'événements et d'états internes situés dans la vie mentale propre du sujet. L'idée à examiner est que l'aperception interne est d'une manière ou d'une autre essentielle à la conscience phénoménale. Pour commencer, un état mental n'est phénoménalement conscient que si son sujet en a une aperception interne. Mais, en un sens plus fort, l'*espèce correcte* d'aperception interne n'est pas seulement une condition nécessaire, mais aussi une condition suffisante de la phénoménalité. C'est l'aperception interne qui, ultimement, rend l'état mental [17] phénoménalement conscient en général. Car c'est quand le sujet en a cette aperception interne que l'état acquiert un caractère subjectif, et le caractère subjectif est ce qui rend un état phénoménalement conscient en général.

On peut trouver douteuse l'idée qu'on apercevrait toujours ses expériences conscientes présentes. Au moment où j'ai mon expérience consciente du ciel bleu, je ne prête pas la moindre attention à moi-même ni à mon expérience. Mon aperception est focalisée sur le ciel bleu, non sur l'expérience que j'en ai. Cependant, le fait que mon aperception n'est pas focalisée sur mon expérience n'implique pas que je n'en ai pas du tout d'aperception. Alors que le focus de notre aperception est pour nous extrêmement manifeste, il y a toute une série d'objets et de caractères qui sont systématiquement rejetés à la *périphérie* de l'aperception. Ainsi, au moment où je regarde fixement l'ordinateur devant moi, l'ordinateur occupe le focus de mon aperception visuelle. Mais on trouve à la périphérie de mon aperception visuelle un grand nombre d'objets situés au côté opposé de mon bureau : un stylo à bille, une tasse de

¹ Cette manière de voir engendre immédiatement un grand nombre de problèmes et d'objections potentielles. Cependant je ne la livre pas ici comme un argument officiel, mais comme un signe en direction d'une conception de nature à motiver l'idée que les états conscients sont nécessairement représentés. L'argumentation « officielle » sera présentée au chapitre 4.

café, une copie du *Tractatus*¹. Comme je le soutiendrai au chapitre 5, la distinction entre focus et périphérie ne vaut pas seulement pour l'aperception visuelle, mais pour toute aperception, y compris l'aperception interne. Ma thèse est que dans le cas de mon expérience du ciel, bien que je n'aperçoive pas focalement mon expérience elle-même, mais le ciel, j'aperçois néanmoins l'expérience elle-même périphériquement. C'est-à-dire que l'expérience combine une aperception externe focale du ciel avec une aperception interne périphérique d'elle-même.

Plausiblement, apercevoir quelque chose, c'est simplement se le représenter de la bonne manière. Apercevoir un arc-en-ciel, pour un sujet, *c'est simplement* abriter en soi l'espèce correcte de représentation mentale de l'arc-en-ciel. Ce traitement représentationnel de l'aperception s'étend également à l'aperception interne. Ainsi, apercevoir, mettons, une anxiété croissante, *c'est simplement* abriter en soi une représentation mentale de son anxiété. Du traitement représentationnel de l'aperception et de la thèse selon laquelle l'aperception interne est essentielle à la conscience phénoménale, il suit que les états conscients sont essentiellement des états dont leur sujet a une représentation mentale.

En résumé, ce qui motive la thèse que les états conscients sont conscients en vertu du fait d'être représentés de la bonne manière vient de [18] l'idée que les états conscients sont des états que nous apercevons, quoique périphériquement, et qu'une telle aperception est une forme de représentation. Nous pouvons ainsi présenter le sous-argument suivant en faveur de la première prémisse de l'argument principal. Pour tout état conscient *C* d'un sujet *S*,

1. *C* est conscient en vertu du fait que *S* aperçoit *C* de manière appropriée ;
2. que *S* aperçoive *C* de manière appropriée, cela veut dire que *C* est représenté par *S* de manière appropriée ; donc
3. *C* est conscient en vertu du fait d'être représenté par *S* de manière appropriée ; donc
4. *C* est conscient en vertu du fait d'être représenté de manière appropriée.

La phase suivante de l'argument principal consiste à passer de « être représenté » à « être autoreprésenté ».

¹ Ces objets se présentent de manière assez vague et imprécise dans mon expérience, mais ils se présentent néanmoins. Ainsi je ne pourrais jamais savoir, sur la base de ma seule expérience, que le livre est le *Tractatus*. Mais il serait évidemment erroné d'en inférer que je n'aperçois le livre en aucune manière. Bien plutôt, nous devrions dire que je l'aperçois, mais *périphériquement* et non *focalement*. Ces problèmes seront discutés plus en détail au chapitre 5.

7. Aperception interne et autoreprésentation

Quand un état conscient est représenté, soit il est représenté par lui-même, soit il est représenté par un état autre que lui, par un état numériquement distinct. Dans la mesure où les états conscients sont conscients précisément en vertu du fait d'être représentés de la bonne manière, il y a trois possibilités : (a) tous les états conscients sont conscients en vertu du fait d'être représentés par eux-mêmes ; (b) tous les états conscients sont conscients en vertu du fait d'être représentés par des états numériquement distincts ; (c) certains états conscients sont conscients en vertu du fait d'être représentés par eux-mêmes et d'autres en vertu du fait d'être représentés par des états numériquement distincts. L'autoreprésentationalisme soutient que (a) est vrai. L'argumentation sera poursuivie au chapitre 4. Je me borne ici à une esquisse préliminaire d'une part centrale de la stratégie argumentative du même chapitre.

Une manière d'argumenter en faveur de (a) ci-dessus est de le faire par élimination — c'est-à-dire en argumentant contre (b) et (c). Il n'est guère plausible, à première vue, que certains états conscients soient conscients en vertu du fait d'être représentés d'une certaine manière alors que d'autres seraient conscients en vertu du fait d'être représentés d'une autre manière. On attendrait une certaine unité sous-jacente, à un degré raisonnable d'abstraction, entre les états conscients. En effet, si les états conscients forment une espèce naturelle, comme c'est probable, ils doivent alors se prévaloir d'une « nature sous-jacente » relativement concrète qui soit commune à eux tous. Ces considérations militent contre (c).

L'argument contre (b) procède par dilemme destructif. Quand un état mental est représenté par un état numériquement distinct, il faut que l'état numériquement [19] distinct soit ou bien *conscient*, ou bien *inconscient*. Considérons le dilemme suivant : tous les états conscients sont-ils conscients en vertu du fait d'être représentés par un état conscient numériquement distinct, ou non ? Si ce n'est pas le cas, alors (b1) tous les états conscients ou certains d'entre eux sont conscients en vertu du fait d'être représentés par des états *inconscients* numériquement distincts. (C'est là, en fait, la conception des théories d'ordre supérieur¹².) Si c'est le cas, alors (b2) tous les états conscients sont conscients en vertu du fait d'être représentés par des états conscients numériquement distincts. L'argument contre (b) consiste à montrer que ni (b1) ni (b2) n'est plausible.

¹ Plus spécialement, les théories d'ordre supérieur contemporaines soutiennent que *la plupart* des états conscients sont conscients en vertu du fait d'être représentés de manière appropriée par des états inconscients, mais que *certaines* sont conscients en vertu du fait d'être représentés de manière appropriée par des états conscients. Les seconds sont les états conscients en train d'être explicitement introspectés. Les premiers sont tous les autres états conscients, et ils constituent évidemment la grande majorité des états conscients.

² Voir, dans le présent recueil, l'introduction ainsi que les textes de D. Rosenthal et P. Carruthers (*N.d.T.*).

L'argumentation contre (b2) est aisée : (b2) mène à une régression à l'infini. Si tout état conscient était nécessairement représenté par un état conscient numériquement distinct, alors l'occurrence d'un seul état conscient impliquerait une infinité d'états mentaux. Mais cela n'est pas plausible à un double titre : on échoue par là à expliquer ce qui rend conscient un état conscient, et ce n'est pas plausible (voire même possible) empiriquement¹.

Quant à (b1), le principal argument qu'on peut lui opposer est un peu plus délicat et je n'en donnerai ici qu'un aperçu². J'ai dit dans la section précédente que les états conscients étaient des états que nous apercevons. Mais comment savons-nous que nous apercevons nos états conscients ? Autrement dit, sur quelle base en venons-nous à souscrire à la thèse suivant laquelle les états conscients sont des états que nous apercevons ? Il semble incorrect de dire que nous souscrivons à cette thèse sur la seule foi de l'évidence de troisième personne. Ce n'est pas comme si nous avions consulté des données expérimentales pointant vers l'existence d'une telle aperception. Bien plutôt, nous y souscrivons sur la foi d'une certaine espèce de connaissance de première main, de première personne, de notre aperception de nos états conscients. Cependant, nous n'aurions pas une telle connaissance de notre aperception si cette aperception était fondée dans des représentations inconscientes. Car notre connaissance des états inconscients est toujours une connaissance à la troisième personne. Ainsi, si notre aperception de nos états conscients était fondée dans des états inconscients, [20] la connaissance que nous en avons devrait être une sorte de connaissance à la troisième personne qu'en fait elle ne semble pas être.

Encore une fois, ce n'est là que l'esquisse d'une ligne argumentative contre (b1) qui sera étoffée et défendue avec beaucoup plus de détails au chapitre 4 du présent ouvrage. Il faut noter, cependant, que puisque (b1) est en fait la conception des théories d'ordre supérieur, un grand nombre des arguments de la littérature contre les théories d'ordre supérieur sont d'application. Ces arguments seront également répétés et/ou développés au chapitre 4. Peut-être le plus connu de ces arguments est-il l'argument des représentations d'ordre supérieur « qui manquent leur cible » [*targetless*]. D'après (b1), ce qui rend conscient un état conscient est qu'il est ciblé par une représentation numériquement distincte, donc d'ordre supérieur. Mais qu'arrive-t-il quand un sujet a une représentation d'ordre supérieur qui représente

¹ Comme le dit Leibniz, si tel devait être le cas, nous ne dépasserions jamais notre première pensée. Il écrit ainsi : « Il n'est pas possible que nous réfléchissions toujours expressément sur toutes nos pensées ; autrement l'esprit ferait réflexion sur chaque réflexion à l'infini. (...) Il faut bien (...) qu'il y ait enfin quelque pensée qu'on laisse passer sans y penser ; autrement on demeurerait toujours sur la même chose. » (G.W. Leibniz, *Nouveaux Essais*, II, 1, cité dans R.J. Gennaro, « Leibniz on consciousness and self-consciousness », dans R.J. Gennaro & C. Huenemann (éds.), *New Essays on the Rationalists*, Oxford, Oxford University Press, 1999, p. 355-356).

² Je le développerai plus en détail au chapitre 4. J'ai d'abord proposé cet argument dans U. Kriegel, « Consciousness as intransitive self-consciousness : Two views and an argument », *Canadian Journal of Philosophy*, 33, 2003, p. 103-132.

trompeusement non seulement les *propriétés* d'une cible d'ordre inférieur, mais jusqu'à son *existence* ? La théorie d'ordre supérieur semble obligée de dire que le sujet n'a pas d'expérience consciente, même si ce qu'il traverse est subjectivement indiscernable de l'accomplissement d'une expérience consciente. Cela veut dire, de façon absurde, qu'il n'y aurait pas d'état conscient où serait le sujet, mais qu'il y aurait un effet que ça fait pour lui d'être dans l'état conscient où il n'est pas¹.

Si les réflexions qui précèdent sont correctes, alors ni (b2) ni (b1) ne peut fonctionner, ce qui, sachant que (c) n'est pas viable, implique que (a) est la seule position viable : tous les états conscients sont conscients en vertu du fait d'être représentés par eux-mêmes. Pour récapituler, l'argument présenté en ce sens gravite autour de deux idées de base. D'abord, il n'est pas plausible qu'un état conscient soit conscient en vertu du fait d'être représenté par un *état inconscient*. Mais s'il est conscient en vertu du fait d'être représenté par un état conscient, l'état conscient représentant ne peut être numériquement distinct de l'état conscient représenté, sous peine de régression ou de manquer d'unité. Il s'ensuit que les états conscients représentant et représenté sont un seul et même état — c'est-à-dire que les états conscients sont autoreprésentants. Nous pouvons ainsi présenter le sous-argument suivant en faveur de la seconde prémisse de l'argument principal. Pour tout état conscient *C*,

1. ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté inconsciemment ; et
2. ce n'est pas le cas que *C* est conscient en vertu du fait d'être consciemment représenté par un état numériquement distinct ; donc [21]
3. ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté par un état numériquement distinct.

Si nous relient l'argument de cette section à l'argument de la section précédente, nous obtenons une version partiellement plus étoffée de l'argument principal. Pour tout état conscient *C* d'un sujet *S*,

1. *C* est conscient en vertu du fait que *S* aperçoit *C* de manière appropriée ;
2. que *S* aperçoive *C* de manière appropriée, cela veut dire que *C* est représenté par *S* de manière appropriée ; donc
3. *C* est conscient en vertu du fait d'être représenté par *S* de manière appropriée ; donc
4. *C* est conscient en vertu du fait d'être représenté de manière appropriée ;
5. ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté inconsciemment ; et
6. ce n'est pas le cas que *C* est conscient en vertu du fait d'être représenté consciemment par un état numériquement distinct ; donc

¹ Une fois encore, je développerai ce point, ainsi que d'autres problèmes liés aux théories d'ordre supérieur, au chapitre 4.

7. ce n'est pas le cas que C est conscient en vertu du fait d'être représenté par un état numériquement distinct ; donc
8. C est conscient en vertu du fait d'être représenté par lui-même de manière appropriée ; c'est-à-dire que
9. C est conscient en vertu du fait de se représenter lui-même de manière appropriée.

L'argument principal peut encore être étoffé, bien entendu. La présente formulation visait principalement à donner une idée de l'attrait de l'approche autoreprésentationnelle de la conscience phénoménale au niveau le plus fondamental.