

Pourquoi le représentationnalisme ne peut pas résoudre le « problème difficile » de la conscience phénoménale

La théorie représentationnaliste de la conscience phénoménale

Par le terme de **conscience phénoménale** (Block, 1995) (ainsi que par ceux d'« effet que cela fait » (Nagel, 1974), de « *qualia* » (Jackson, 1982) et d'« expérience consciente »), l'on désigne la dimension expérientielle et vécue de la vie mentale. Cette « conscience phénoménale » nourrit depuis plus de trente ans un grand nombre de discussions en philosophie de l'esprit, dans la mesure où elle semble poser un problème fondamental au projet d'explication physique de l'esprit – que David Chalmers (Chalmers, 1995) a nommé le « problème difficile » (*hard problem*) de la conscience. Ce dernier s'oppose aux problèmes « faciles », que tentent d'ores et déjà de traiter les neurosciences et les sciences cognitives, et peut se subdiviser en deux sous questions (Chalmers, 1996) : pourquoi et comment un ensemble donné de processus cérébraux s'accompagne-t-il d'une expérience phénoménale d'une nature particulière, et pas d'une autre nature – par exemple, d'une expérience de rouge plutôt que d'une expérience de vert ? (problème du **contenu phénoménal**). D'autre part, pourquoi et comment un ensemble donné de processus cérébraux s'accompagne-t-il d'une expérience phénoménale, quelle qu'elle soit ? (problème de **l'existence d'une phénoménalité**).

Une des approches les plus discutées de cette forme de conscience tente d'analyser celle-ci à travers la notion de « **représentation** ». Cette approche « **représentationnaliste** », comprise à un haut niveau de généralité, s'appuie sur l'idée selon laquelle avoir une expérience consciente du monde implique toujours de représenter le monde (l'environnement externe, ou mon propre corps) d'une certaine façon, c'est-à-dire de lui attribuer certains traits. Et tout changement dans mon expérience phénoménale semble impliquer un changement de la façon dont je représente le monde : ainsi, si j'ai l'expérience visuelle d'un objet A doté d'une forme X, je ne peux pas imaginer que la forme dont je fais l'expérience change sans que change du même coup la façon dont je représente cet objet, c'est-à-dire sans que change les traits que ma perception attribue à cet objet (Siewert, 1998, chap. 7). Il existe toutefois des formes de représentationnalisme (Dretske, 1995, 2003; Tye, 1995, 2000) qui ne se contentent pas de remarquer le lien essentiel entre expérience consciente et représentation, mais utilisent la notion de représentation pour **expliquer** et **réduire** la conscience phénoménale. De cette manière, elles prétendent résoudre le « problème difficile » : l'expérience consciente ne consistant en rien d'autre qu'en processus représentationnels, le mystère apparemment inhérent à celle-ci disparaît. Ce sont ces théories que nous allons critiquer, en soulevant à leur encontre une objection qui nous semble particulièrement fondamentale. Ces théories sont les **théories représentationnalistes fortes et réductives de la conscience phénoménale**. Elles sont fortes en ce qu'elles posent qu'avoir une expérience consciente dotée d'un contenu phénoménal X, ce n'est rien d'autre que représenter (d'une certaine manière) le monde comme possédant un trait X. Elles sont réductives en ce qu'elles supposent que le processus représentationnel peut être entièrement naturalisé, c'est-à-dire explicité en termes physiques – supposition qui est essentielle pour qu'elle puisse prétendre à la résolution du problème difficile.

Cette explication de la représentation en termes physiques n'est pas évidente. En effet, une « représentation » a des propriétés singulières : par exemple, elle peut être « correcte » ou « incorrecte » (elle peut représenter un objet de manière véridique ou fautive, et peut même représenter quelque chose qui n'existe pas), ce qui ne semble pas constituer une caractéristique classique des entités matérielles – qui, elles, se contentent d'exister sans pouvoir être jugées sous

l'angle de la correction. Des philosophes parmi lesquels Fred Dretske¹ ont toutefois développé une théorie naturelle de la représentation au cours des années 1980. Celle-ci fait fond sur la notion d'indicateur naturel : un indicateur est un élément matériel qui indique autre chose que lui en raison d'une covariation non-accidentelle – c'est-à-dire, typiquement, d'un lien de nature causale. Par exemple, les cernes d'un arbre « indiquent » naturellement l'âge de cet arbre. Toutefois, l'indication n'est pas encore la représentation, puisqu'elle suppose un rapport réel entre la chose indiquée et l'indicateur, et ne peut en aucun cas être incorrecte : l'indication ne se « trompe » pas, elle est ou elle n'est pas. Pour construire la notion de représentation à partir de celle d'indication, la notion de « fonction » est exigée. L'idée est que X représente Y si et seulement si X a pour fonction d'indiquer Y. En effet, le fait que X possède cette fonction fait que, si à un instant donné X ne covarie plus avec Y, il n'y a pas simplement absence d'indication mais bel et bien représentation incorrecte. On peut donc dire que l'état X d'un système S représente Y s'il a pour fonction d'indiquer Y. Une dernière difficulté se pose alors à la théorie naturelle de la représentation : comment peut-il exister une fonction objective, naturelle ? Ne faut-il pas d'emblée un esprit humain pour attribuer, par convention, une fonction à un système ? L'idée de Dretske, c'est qu'une chose peut acquérir une fonction naturelle grâce au processus de sélection naturelle et à l'histoire évolutionnaire : un mécanisme X a par exemple pour fonction naturelle d'indiquer Y s'il a été sélectionné justement pour cette raison au cours de l'évolution. Il existe donc, selon Dretske, en plus des représentations conventionnelles, de véritables représentations naturelles, qui prennent place dans des systèmes physiques sans nécessiter l'interprétation d'un esprit préexistant, et qui sont définies de la façon suivante : **X représente Y de façon naturelle si et seulement si X a été sélectionné au cours de l'évolution naturelle parce qu'il indique Y en raison d'une covariation non-accidentelle.**

La théorie représentationnaliste pose donc que l'on peut réduire la conscience phénoménale à la représentation naturelle ainsi comprise. Elle peut se résumer en deux thèses principales :

Thèse 1 : Être conscient phénoménalement d'un objet ou d'une propriété, ce n'est rien d'autre que représenter naturellement (et sous certaines conditions²) cet objet ou cette propriété.

Thèse 2 : Le contenu phénoménal d'un état (les propriétés qualitatives vécues d'un état mental) n'est rien d'autre que le contenu représentationnel de cet état, c'est-à-dire que les propriétés que cet état représente.

Les arguments en faveur du représentationnalisme sont nombreux et parfois techniques. On justifie souvent le représentationnalisme en s'appuyant sur la thèse phénoménologique de la « transparence » de l'expérience perceptive (Tye, 1995, 2000) ou sur le caractère intensionnel du contexte propositionnel propre aux jugements phénoménaux (Tye, 1995). Nous n'avons pas ici le temps de développer ces arguments célèbres, mais nous désirons attirer l'attention sur une caractéristique du représentationnalisme qui le rend particulièrement séduisant : sa capacité à résoudre le problème classique de la localisation des propriétés phénoménales (Dretske, 1995, 2003). En effet, quand j'ai une expérience perceptive, par exemple de bleu, la question se pose de savoir où se trouve le bleu dont je fais l'expérience. Est-il situé au-dehors, dans le monde ? Il semble que non, puisque je peux me contenter d'halluciner ce bleu. Est-il dans mon cerveau ? Non, car il n'y a rien qui soit phénoménalement bleu dans ma boîte crânienne. Le philosophe semble alors contraint de situer le bleu phénoménal dans une dimension purement mentale, non-physique : ce qui d'une part oblige à rompre avec le matérialisme et d'autre part débouche sur des théories de type « sense-data », avec leur cortège de difficultés. Le représentationnalisme, en revanche, donne à cette question une réponse à la fois claire et compatible avec le matérialisme : le bleu dont je fais l'expérience se situe dans le monde, mais pas dans le monde extérieur : il existe dans le monde

¹ Il existe de nombreuses théories naturalistes de la représentation. Pour des raisons de commodité, nous nous limiterons ici à une version (assez simplifiée) de la théorie de Dretske (Dretske, 1981, 1995), qui a l'avantage de figurer parmi les plus synthétiques.

² La suite de notre article se penchera précisément sur le problème de la spécification de ces conditions.

représenté par mon état perceptif, **dans le monde que mon état perceptif vise et « décrit »**. On comprend alors en quoi le représentationnalisme est séduisant : il est compatible avec le matérialisme sans pour autant identifier l'esprit et le cerveau³. Le cerveau est en effet le lieu des véhicules représentationnels, c'est-à-dire de la réalisation physique des états représentationnels, tandis que l'esprit est le domaine des contenus représentés et visés par ces états. Cette solution satisfait les exigences du matérialisme sans mener aux perplexités qui naissent d'ordinaire de l'identification psycho-physique⁴.

Une objection à la théorie représentationnaliste

De nombreuses objections ont été adressées aux théories représentationnalistes de la conscience phénoménale. Généralement, ces objections consistent en contre-exemples à la théorie, présentant des cas réels (Peacocke, 1983) ou imaginaires (Block, 1990; Levine, 2003) – le cas le plus célèbre étant celui du spectre inversé (Shoemaker, 1984) – de non-coïncidence entre le contenu phénoménal d'un état mental et son contenu représentationnel. On le voit, ce type d'objection se focalise autour du problème du « contenu », qui correspond à la « thèse 2 » du représentationnalisme. Toutefois, comme nous l'avons noté, le « problème difficile » de la conscience ne se réduit pas au problème du contenu phénoménal : en réalité, le **problème de l'existence d'une phénoménalité** possède une certaine primauté.

Une objection au représentationnalisme qui fait fond sur ce problème de l'existence de la phénoménalité a été soulevée et nous apparaît comme particulièrement fondamentale : c'est « **l'objection de la démarcation** » (Byrne, 2001; Kriegel, 2002; Seager, 2003; Seager & Bourget, 2007; Stoljar, 2007)⁵. Elle prend pour point de départ la remarque selon laquelle, si le représentationnalisme prétend expliquer la conscience phénoménale, il doit être en mesure de justifier et d'expliquer la différence (la « démarcation ») entre états mentaux phénoménaux et non-phénoménaux. L'objection consiste à montrer que les théories représentationnalistes n'y parviennent pas.

Le représentationnaliste pourrait tout d'abord soutenir que la démarcation entre les états phénoménaux et les états non-phénoménaux correspond à la différence entre les états qui sont des représentations naturelles et les autres. Toute représentation naturelle, en ce sens, serait automatiquement phénoménale ; et le contenu d'une telle représentation serait un contenu phénoménal. C'est la thèse, élégante mais risquée, que semble soutenir Fred Dretske⁶. Le problème est qu'une telle thèse est fortement contre-intuitive : si en effet on pose qu'*X* représente naturellement *Y* si *X* a pour fonction naturelle d'indiquer *Y*, alors il existe beaucoup de représentations naturelles en dehors de notre cerveau (par exemple, le système digestif ou encore

³ Le représentationnalisme présente aussi l'avantage de donner une image de l'esprit unifiée et cohérente avec les théories actuelles des sciences cognitives. Le concept de « représentation » est en effet souvent considéré (Fodor, 1981) comme la clé permettant de comprendre le fonctionnement cognitif de l'esprit. Le représentationnalisme de la conscience phénoménale, qui soutient que le même concept est suffisant pour expliquer la dimension expérientielle de l'esprit, porte donc une promesse d'unité.

⁴ On voit toutefois que, étant donné que seul le véhicule de la représentation est véritablement matériel, il est seul à posséder une véritable efficace causale : l'épiphénoménalisme du contenu représentationnel (et donc du contenu phénoménal) menace donc le représentationnalisme. Nous ne pouvons pas traiter ici de ce problème, auquel Dretske a par ailleurs tenté de répondre (Dretske, 1988).

⁵ Ces auteurs évoquent ou soutiennent cette objection de manières distinctes. J'en propose ici une reformulation relativement originale.

⁶ Dretske (Dretske, 1995, p. 19) fait une remarque qui semble aller contre ce point et rapprocherait sa théorie de celle de Tye, mais cet aspect disparaît ensuite (Dretske, 2003).

l'épiderme des humains manifestent de nombreux paramètres représentationnels). Peu d'entre nous désirent toutefois affirmer que notre estomac et notre peau ont des expériences conscientes.

Si le représentationnaliste refuse d'avaler une telle couleuvre, il doit affirmer que certaines représentations naturelles ne sont pas phénoménales, et donc fournir un critère de démarcation satisfaisant pour les distinguer de celles qui le sont. Michael Tye (Tye, 1995, 2000) en propose un, en affirmant que le contenu phénoménal d'un état s'identifie au « **PANIC** » de ce même état (PANIC étant un acronyme pour : Poised Abstract Non-conceptual Intentional Content⁷) : seuls les contenus PANIC sont phénoménaux. Dans cette caractérisation, c'est sur la condition « **disponible** » que semble reposer principalement la démarcation. Cette condition signifie en effet qu'un contenu représentationnel donné doit être disponible pour influencer le système cognitif central (notamment le système de formation des croyances et des désirs), afin d'être phénoménalement conscient. Ainsi peuvent être mis de côté les contre-exemples précédemment évoqués : les représentations du système digestif ne sont pas disponibles pour un centre cognitif quelconque, à la différence de nos représentations perceptives, ce qui explique qu'elles ne donnent pas lieu à des expériences conscientes.

Cette proposition de Tye peut toutefois être critiquée selon trois perspectives. Tout d'abord (1), il semble que certaines situations perceptives particulières, comme les perceptions subliminales ou encore les perceptions qui concernent les patients *blindsight*⁸, peuvent constituer des contre-exemples. Dans le cas d'une représentation *blindsight* par exemple, il existe un état doté d'un contenu représentationnel, abstrait, non-conceptuel, qui vient (dans certaines situations de choix forcé) influencer le système central de formation des croyances ; pourtant, d'après les sujets eux-mêmes, ce contenu n'est pas phénoménal. La réponse de Tye est que, dans ce cas, la condition « disponible » n'est remplie qu'en apparence, et que le contenu intentionnel **n'est pas disponible de la bonne manière** (Tye, 2000), notamment dans la mesure où les perceptions *blindsight* ne donnent pas lieu à de véritables « croyances » chez les sujets, mais que ces derniers se contentent de « deviner » le contenu représentationnel de leurs états perceptifs⁹. Le problème est qu'un *superblindsight*, capable de produire des états perceptifs formant la base de véritables croyances, sans pour autant que ceux-ci soient plus phénoménaux que les états *blindsight* classiques, semble concevable. Une telle concevabilité constitue un premier argument contre l'idée que le critère de « disponibilité » suffit à assurer la démarcation.

Par ailleurs (2), toujours sur le plan des contre-exemples imaginaires, il semblerait possible d'imaginer un dispositif physique capable de « bloquer » l'influence d'un état représentationnel sur le système cognitif central – par exemple en bloquant certaines interactions neuronales. Or, les états représentationnels doivent être « disponibles » pour influencer ce système, afin d'être phénoménaux : cela signifie qu'ils ne l'influencent pas toujours effectivement. Il devrait donc être possible d'imaginer (Seager, 2003) un dispositif physique particulièrement sophistiqué, qui se contente de bloquer l'influence sur le système cognitif central de ceux des états représentationnels qui sont en théorie disponibles mais qui, en fait, n'ont aucune action sur ce système. Un tel mécanisme serait donc « purement contre-factuel » et dénué d'effet concret : pourtant, il annulerait la « disponibilité » des états représentationnels concernés, ce qui signifie qu'il les rendrait non-

⁷ Que l'on peut traduire par : « Contenu Intentionnel, Non conceptuel, Abstrait et Disponible ». Dans ce contexte, « intentionnel » peut être vu comme synonyme de « représentationnel ».

⁸ A la suite d'une lésion du cortex visuel, certains patients ont une zone de leur champ visuel qui devient « aveugle » : ils ne se déclarent visuellement conscients de rien à propos de cette zone. Toutefois, sous certaines conditions, ils parviennent à accéder à certaines informations visuelles dans cette zone.

⁹ En s'appuyant sur les travaux récents de Tye (Tye, 2009), on pourrait formuler une réponse encore plus précise : un état représentationnel doit, pour être phénoménal, être disponible pour contribuer à la formation et à la justification de connaissance. Toutefois, la concevabilité d'un *superblindsight* qui permet de remplir les conditions reste envisageable, quoique sans doute plus contestable.

phénoménaux. On voit toutefois mal en quoi la simple **présence** d'un tel mécanisme, dépourvu d'effet réel sur le fonctionnement du système cognitif, pourrait annuler la phénoménalité de certains états. Ce qui constitue un argument contre l'idée selon laquelle la « disponibilité » pourrait jouer un véritable rôle de démarcation.

Or (3), la réponse de Tye (Tye, 2003) à ce dernier contre-exemple repose sur l'idée qu'un tel dispositif hypothétique n'annulerait pas la « disponibilité » des états représentationnels concernés, car une telle disponibilité leur est intrinsèque (de même que le fait de protéger un verre de cristal avec du papier bulle n'annule pas sa fragilité). Mais cette réponse de Tye montre en réalité les difficultés conceptuelles inextricables que rencontre celui qui voudrait faire reposer le principe de la démarcation sur ce critère de disponibilité. En effet, « être disponible » est une propriété dispositionnelle, que l'on peut interpréter soit de manière relationnelle (comme la propriété « être accessible par l'autoroute » ou « être joignable par téléphone »), soit de manière intrinsèque (comme « être fragile », que nous venons d'évoquer). Dans le premier cas, la disponibilité d'un état représentationnel est annulé par un dispositif du type de celui imaginé par Seager (tout comme le fait de me situer dans une cage de Faraday fait disparaître ma propriété « d'être joignable par téléphone », même si personne ne m'appelle), ce qui porte un coup à la thèse de Tye : car on voit mal comment un phénomène occurrent, tel que l'expérience consciente, pourrait être annulé par la présence d'un mécanisme sans effet – par l'existence d'une disposition qui ne s'actualise jamais. Dans le second cas, qui est l'interprétation sur laquelle Tye s'appuie pour répondre à Seager, le cœur de la propriété de « disponibilité » d'un état représentationnel ne concerne pas son rapport potentiel à un centre cognitif quelconque, mais consiste en un ensemble de propriétés intrinsèques qui sont la cause, dans certaines conditions particulières, de ce rapport potentiel à un centre cognitif. Autrement dit, cette seconde interprétation de la « disponibilité » la définit comme l'ensemble des propriétés intrinsèques de l'état représentationnel qui réalisent physiquement la propriété fonctionnelle et relationnelle d'interaction potentielle avec le centre cognitif – qui réalisent la « disponibilité » de la première interprétation.

Mais dans ce cas également, la théorie de Tye se voit contrecarrée. Car si la simple présence de ces propriétés réalisatrices doit pouvoir jouer le rôle de démarcation, alors ce sont ces propriétés elles-mêmes qui doivent se voir explicitées et détaillées – et pas seulement sous l'angle de leurs effets dans des situations typiques, ce qui est pourtant le cas si on les désigne par le mot de « disponibilité ». Le problème, c'est que nous ne savons pas comment de telles propriétés réalisatrices – qui ne sont pas des propriétés représentationnelles, ni fonctionnelles, mais intrinsèques et physiques – peuvent justifier la démarcation et expliquer la présence de l'expérience conscientes. Faire reposer la démarcation sur la présence de certaines propriétés physiques spécifiques ne fait que reposer le problème corps-esprit dans sa forme la plus classique. En ce sens, si le représentationnalisme doit s'appuyer sur un critère de « disponibilité » ainsi interprété pour fonder la démarcation, il se voit rattrapé par le « problème difficile » tel qu'il s'adresse aux théories de l'identité psycho-physique.

Cette objection de la démarcation que nous venons de soulever permet de mettre en lumière l'une des faiblesses les plus cruciales du représentationnalisme : ce dernier ne parvient pas à justifier de manière satisfaisante (non *ad hoc*) la **démarcation** entre les contenus phénoménaux et non-phénoménaux, justification qui semble pourtant requise par la résolution du problème de **l'existence même d'une phénoménalité**. En ce sens, même les tentatives les plus ambitieuses et les plus achevées de la théorie représentationnaliste de la conscience phénoménale semblent être reconduites au « problème difficile » de la conscience.

Bibliographie

Block, N. (1990). Inverted earth. *Philosophical Perspectives*, 4, 53-79.

- Block, N. (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences*, 18, 227–247.
- Byrne, A. (2001). Intentionalism defended. *Philosophical Review*, 110(2), 199-240.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-19.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge (Mass.): MIT Press.
- Dretske, F. (1988). *Explaining Behavior : Reasons in a World of Causes*. MIT Press.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Dretske, F. (2003). Experience as representation. *Philosophical Issues*, 13(1), 67-82.
- Fodor, J. (1981). *Representations*. Cambridge, Massachusetts: MIT Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(April), 127-136.
- Kriegel, U. (2002). PANIC theory and the prospects for a representational theory of phenomenal consciousness. *Philosophical Psychology*, 15(1), 55-64.
- Levine, J. (2003). Experience and representation. Dans Q. Smith & A. Jokic (Éd.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(October), 435-50.
- Peacocke, C. (1983). *Sense and Content: Experience, Thought, and Their Relations*. Oxford: Oxford University Press.
- Seager, W. (2003). Tye on consciousness : time to panic? *Philosophical Studies*, 113(3), 237-247.
- Seager, W., & Bourget, D. (2007). Representationalism about consciousness. Dans M. Velmans & S. Schneider (Éd.), *The Blackwell Companion to Consciousness*. Oxford: Blackwell Publishing.
- Shoemaker, S. (1984). The inverted spectrum. *Identity, Cause and Mind*. Cambridge: Cambridge University Press.
- Siewert, C. (1998). *The Significance of Consciousness*. Princeton: Princeton University Press.
- Stoljar, D. (2007). Consequences of intentionalism. *Erkenntnis (Special Issue)*, 66(1-2), 247--70.
- Tye, M. (1995). *Ten problems of consciousness : a representational theory of the phenomenal mind*. Representation and mind (Vol. 1-1). Cambridge (Mass.): the MIT Press.
- Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.
- Tye, M. (2003). On the virtue of being poised – reply to Seager. *Philosophical Studies*, 113(3), 275-280.
- Tye, M. (2009). *Consciousness Revisited : Materialism without Phenomenal Concepts*. Cambridge (Mass): MIT Press.