



HAL
open science

Lorsque je ne pense pas ce que je pense: l'infailibilité à l'épreuve de la pureté neurale?

David Roman Zarebski

► **To cite this version:**

David Roman Zarebski. Lorsque je ne pense pas ce que je pense: l'infailibilité à l'épreuve de la pureté neurale?. *Revue Etudiante de Philosophie Analytique*, 2013, 7, pp.37-45. halshs-01136578

HAL Id: halshs-01136578

<https://shs.hal.science/halshs-01136578>

Submitted on 18 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

**LORSQUE JE NE PENSE PAS CE QUE JE PENSE:
L'INFAILLIBILITÉ À L'ÉPREUVE DE L'ARGUMENT
DE LA PURETÉ NEURALE?**

David ZAREBSKI
(Paris I - IHPST)

L'idée que je suis, sujet, réflexivement conscient de ce dont je suis conscient est certainement l'une des intuitions les plus spontanées qui accompagne notre vie mentale. D'une manière similaire à nombre d'intuitions spontanées que la conscience peut entretenir vis-à-vis d'elle-même, cette accessibilité complète et inconditionnelle de la conscience a connu nombre de raffinements aussi bien empiriques que conceptuels en vue de mener son étude dans la sure voie de la science. L'argument dit *de la pureté neurale* (APN), à l'analyse duquel nous consacrerons cet article, dépasse cependant de loin l'enjeu d'une accessibilité complète – telle que dans un débat qui animerait Phénoménologues et Psychologues – en cela que, biproduit d'une opposition scientifique majeure dans le paysage scientifique contemporain, se joue à travers cet argument la possibilité d'une expérience phénoménale pleine et entière sans conscience aucune de cette dernière en première personne. Nous entendons opposer à la possibilité d'une telle expérience soit, *in fine*, à l'APN, un contre-argument non-introspectif jusqu'alors négligé par la littérature. Après une présentation de la logique interne de l'APN, de son contexte d'apparition et de ses implications pour la possibilité d'une étude scientifique de la conscience, nous montrerons de quelle manière la stratégie commune des contre-arguments classiques entrave leur portée contre cet argument. Ces deux points menés à bien, nous nous emploierons à esquisser une voie argumentative basée sur le rôle causal des régions supérieures sur les régions visuelles jusqu'alors absente du débat.

1) Émergence et logique interne de l'argument de la pureté neurale

Nous introduirons le contexte d'émergence de cet argument avant de le formuler ou détailler sa logique interne. Celui-ci naît originellement d'une question méthodologique relative à l'investigation

scientifique de la Conscience: celle de pouvoir attester, en troisième personne, du fait qu'un certain sujet S perçoit consciemment une certaine propriété x ($S(Px)$). Deux théories s'affrontent directement sur cette problématique. D'un côté de la table, l'on peut considérer avec les Théories Basées sur la Récurrence (*Recurrence-Based Theories* ou RBT) – voir (Lamme 2004) – que le corrélat de nature phénoménale de la conscience visuelle est à chercher du côté de boucles locales situées dans les régions ventrales du système visuel. Plus qu'acceptée, cette théorie constitue l'une des applications les plus directes de la distinction entre une conscience phénoménale et une conscience d'accès avancée par Block (1995; 2005).

Le grand modèle rival, la théorie de l'Espace de Travail Global (*Global Workspace Model* ou GWM) défendue par, notamment, (Dehaene, Changeux, and Kerszberg 1998; Dehaene and Changeux 2011), pourrait être qualifiée, *a contrario*, de thèse holiste d'inspiration plus dennettienne¹ en cela qu'elle récuse la distinction blockéenne en cherchant dans les connexions longue distance avec les régions pré-frontales le corrélat physiologique de notre phénoménalité. Les positions épistémologiques de ces deux théories de la conscience vis-à-vis de la problématique méthodologique précédemment évoquée dépendent en grande part de leurs conceptions respectives de la nature de la conscience indépendamment de notre investigation de cette dernière. Un exemple devrait suggérer l'opposition incommensurable qui anime ces théories.

La mémoire iconique présentée en (Sperling, 1960) peut illustrer cette dernière. La présentation rapide (50ms) de trois lignes de quatre lettres ne permet pas au sujet de rapporter plus de quatre de ces douze lettres. Cependant, si l'on indique au sujet trois fréquences de sons correspondant chacune à une ligne particulière avec pour instruction de rapporter la ligne correspondante, il se produit le phénomène intéressant que, quand bien même le son est-il joué 50 ms après la disparition des lettres de l'écran, le sujet est cependant capable de rapporter les quatre lettres de la ligne en question. Pour le dire autrement, ces résultats impliquent que la trace neurale de ces douze lettres reste potentiellement réactivable après coup. Si cette présentation parcimonieuse laisse la

¹ Dennett accorda très tôt au GWM une place privilégiée comme en (Dennett, 2005, p131-136).

délicate question de savoir ce que le sujet perçoit consciemment de côté, l'affaire se complique sitôt que celle-ci émerge en cela qu'une RBT considérera que le sujet était phénoménalement conscient des douze lettres mais ne pouvait accéder qu'à quatre d'entre-elles là où un GWM conclura que le simple fait que le sujet n'ait pas été en mesure de rapporter plus de quatre lettres implique qu'il n'a perçu consciemment que ces quatre-là². Nous nous apprêtons à suggérer que cette question est empiriquement indécidable.

Transposée en des termes fonctionnels, cette impossibilité revient à dire que même une expérience parfaite (*perfect experiment*) ne pourrait, pour des raisons logiques, trancher cette question. Ainsi l'on peut imaginer avec Cohen et Dennett (2011) que la région cérébrale responsable de la discrimination de la couleur rouge se trouve chirurgicalement isolée des régions supérieures et, notamment, pré-frontales. Dans ce contexte, il devient tout à fait envisageable que la région en question s'active en présence d'une pomme rouge sans que le sujet ne puisse accorder à l'expérimentateur qu'il perçoit effectivement la couleur rouge. Cependant, de par la dissociation qu'admet le partisan des RBT, celui-ci peut, sur la seule base de cette information physiologique, affirmer que le sujet perçoit consciemment (conscience phénoménale) cette couleur sans que cette information lui soit accessible. En d'autres termes, $S(Px)$ étant instable en principe, la RBT recourt à l'APN qui, formulé, revient à avancer que:

l'existence d'une correspondance une à une des propriétés visuelles avec leurs corrélats neuraux constitue autant d'index univoques de la perception consciente de ces propriétés dispensant de toute autre manifestation comportementale de cet état mental.

2) *Impuissance des contre-arguments classiques*

Nous entendons montrer dans cette section qu'une propriété commune des contre-arguments des partisans du GWM mène ces derniers à l'alternative inconfortable d'un béhaviorisme ou d'une forme plus

² Cela ne veut pas dire que les autres n'aient pas été perçues. Le GWM laisse la place à l'idée que, en fonction du renforcement généré par le son, les différentes lettres puissent être reconnues comme des lettres sans reconnaissance de chaque symbole individuellement (voir (Kouider et al., 2010)).

globale d'APN.

Une première famille d'arguments consiste à avancer que les partisans des RBT confondent la conscience phénoménale avec l'existence de processus inconscients. Ainsi, outre le *blindsight* et de nombreux phénomènes similaires, il se pourrait tout à fait que le sujet puisse, au moyen d'une série de choix binaires dans une variante de l'expérience Sperling, deviner le reste des douze lettres; rien n'indique que cette performance implique quelque perception consciente de ces autres lettres plutôt qu'une perception subliminale.

En plus de critères méthodologiques – nécessité d'au moins un rapport introspectif pour identifier l'index d'une propriété phénoménale, critère de scientificité, etc – une seconde famille d'arguments repose sur le rôle causal des contenus phénoménaux. Voir du rouge n'est pas une tâche dans quelque Théâtre Intérieur mais l'usage, par un système cognitif, d'une certaine information de l'environnement. Cette approche fonctionnelle de la phénoménalité est aussi bien opposée à l'APN (Cohen & Dennett, 2011) qu'à l'argument connexe dit du débordement phénoménal (*overflow argument*) en vertu duquel le contenu phénoménal dépasserait quantitativement ce qui est susceptible d'être rapporté (Kouider & al., 2010, pp 301-302). Cette stratégie présente cependant le risque d'une forme de béhaviorisme dans la mesure où ces contenus phénoménaux sont alors instanciés pas leurs effets typiques sur le comportement et, notamment, la capacité de S à énoncer ses propres états mentaux.

L'argumentaire de Kouider et al. (2010) se veut cependant moins catégorique en cela qu'il serait déraisonnable d'attendre du sujet le moindre rapport introspectif sur des propriétés de bas niveau dénuées de contre-parties linguistiques. Je suis à cet égard tout à fait capable de percevoir et différencier une certaine nuance de bleu ou un certain angle sans être capable de nommer ces derniers. Dans ce contexte, d'autres tâches, motrices notamment, sont susceptibles de tenir lieu de rapport. Toutefois, si l'on garde à l'esprit la prégnance de mécanismes inconscients de détection postulés par le GWM³, le fait de considérer qu'une certaine

³. Voir en particulier (Dehaene et al. 2006) pour une clarification de l'architecture des niveaux de conscience dans le GWM.

réponse motrice atteste de la perception consciente d'une certaine propriété, loin de congédier l'APN, produit un nouvel APN à l'échelle, cette fois, du système visuo-moteur. S aurait tout aussi bien pu presser le bouton sans percevoir consciemment le stimulus (perception subliminale). Outre lui demander, exclu dans ce contexte, le seul moyen d'attester de $S(Px)$ consiste alors à s'en remettre à un putatif index neural (visuo-moteur, cette fois) de $S(Px)$ soit à étendre l'APN plutôt que l'évacuer.

3) *Approche contraposée*

Nous défendrons l'idée que cette aporie n'est pas tant la conséquence d'une certaine vue de la conscience ou de la nature des contenus phénoménaux que le résultat d'une stratégie exclusivement basée sur une direction de la causalité: des contenus phénoménaux vers le reste de l'architecture cognitive à la manière du schéma suivant.

(a) SI perception(x) ALORS effet sur un autre module

Or, pour peu que l'on prenne acte d'un des biproduits de l'APN, il devrait être possible de se déprendre de l'argument sans tomber sur l'un des écueils précédemment évoqués. En effet, l'APN présente la particularité, de par sa pensée en deux temps, d'induire que l'expérience perceptive et consciente du sujet de l'expérience parfaite serait qualitativement indistinguable d'une autre, entendu que l'accessibilité ne change rien à la phénoménalité. Ainsi que nous l'avons présenté, le GWM est davantage holiste, raison pour laquelle il considère que, par essence, les contenus phénoménaux possèdent non seulement un rôle causal essentiel sur le reste du système mais également que ce rôle causal constitue une condition nécessaire de phénoménalité; idée dennettienne selon laquelle un contenu phénoménal qui serait dénué de fonction constitue une illusion cognitive. Cependant, il convient de distinguer cette conception sur la nature causale des contenus phénoménaux (ratio essendi) de la question de savoir comment attester de leur occurrence (ratio cognoscendi). Le problème du schéma (a) tient dans le fait que les deux régimes se confondent dans l'argumentaire du GWM.

Il existe cependant un moyen simple, si ce n'est d'attester de la nature du x que S percevrait consciemment, du moins de saper l'intuition

de localité ou d'encapsulation qui accompagnent l'APN à sa base en prenant la contraposée du précédent schéma:

(b) Si non(autres modules) ALORS non-id(perception)

Le schéma (b) affirme, littéralement, qu'en l'absence du reste de l'architecture cognitive, le sujet ne peut percevoir de la même manière qu'il le ferait en présence de ces modules. Nombre d'axes de recherches contemporains explorent ce type de dépendance en montrant, à titre d'illustration, que la volonté d'agir, plus que l'action, accélère la détection de propriétés de bas niveau comme l'orientation (Bekking and Neggers 2002), que les basses fréquences spatiales influencent la reconnaissance d'objets (Bar et al. 2006) ou encore que les régions pré-frontales modulent la capacité à discriminer certaines formes simples au milieu de scènes complexes (Barcelo, Suwazono, and Knight 2000).

L'argument peut paraître évident tant l'influence des *processus top-down* sur la visibilité ont été documentés. L'on peut cependant s'étonner que cette voie argumentative n'ait jamais été empruntée dans le contexte du débat opposant les partisans du GWM à ceux des RBT. Cette piste est cependant pertinente dans le débat ainsi que cette variante de l'expérience de pensée de (Cohen and Dennett 2011) devrait le suggérer.

Repartons du sujet chirurgicalement modifié décrit plus tôt. Dans la situation où l'on présentait du rouge, que la région concernée s'activait et que ce dernier ne pouvait rapporter aucune expérience de rouge, le défenseur des RBT pouvait avancer l'APN. Cependant, si un mécanisme top-down similaire à ceux évoqués précédemment nous permettait d'affirmer que les régions frontales modulent cette activité – permettent, par exemple, de distinguer le rouge de l'orange –, suggérant que quand bien même l'on croirait que ces boucles locales constituent le corrélat de notre expérience phénoménale, cette expérience ne pourrait être la même, le partisan des RBT serait alors confronté à une alternative inconfortable.

Il pourrait, premièrement, affirmer que l'expérience phénoménale n'en demeurerait pas moins la même, que cette modulation ne concerne que l'accès à cette information phénoménale mais serrait alors obligé de concéder que la conscience d'accès serait, à sa manière, productrice de contenus phénoménaux puisqu'à même de les modifier après coup. Pourquoi distinguer deux consciences si la seconde ne fait que remplacer le travail de la première?

Il pourrait, également, concéder que l'expérience de couleur serait alors différente mais ne pourrait tout simplement pas prouver qu'elle ait quoi que ce soit à voir avec la perception d'un rouge plutôt que, mettons, le goût de la pistache. En vérité, il n'y aurait pas plus de raison de considérer la région en question responsable de la discrimination des couleurs que de rechercher le corrélat de l'expérience phénoménale dans les régions les plus périphériques du cortex visuel primaire. Dans un cas comme dans l'autre, le *processus top-down* constitue un défi que l'appareil conceptuel des RBT ne peut que difficilement relever.

Conclusion

Défendant une conception de la conscience proche du GWM – holisme, rôle causal des contenus, etc. – nous avons montré que, de par leur focalisation sur le rôle causal des contenus phénoménaux sur les autres modules, les contre-arguments à l'APN ne pouvaient venir à bout de l'intuition à son origine sans basculer dans:

i) une forme de béhaviorisme

ou

ii) une variante à plus large échelle de l'APN.

Là où l'approche causale classique d'un partisan du GWM veut pouvoir identifier la nature de la propriété phénoménale dont S serait conscient, l'approche contraposée est plus modeste en cela qu'elle se contente de montrer l'impossibilité de penser que le sujet de l'expérience parfaite ou un cortex visuel plongé dans une cuve pourraient éprouver la même expérience phénoménale qu'un cerveau entier. L'approche est cependant suffisamment puissante pour rendre impensable une expérience phénoménale dans de semblables cas ; renvoyant alors les RBT à la rigidité de leur dissociation et à l'alternative inconfortable entre:

i) l'abandonner

ou

ii) d'indexer des traitements fonctionnels sur des régions arbitraires sans indice probant pour le faire.

Bibliographie

- Barcelo, F., Suwazono, S. & R. Knight (2000) - Prefrontal Modulation of Visual Processing in Humans (*Nature Neuroscience*, vol.3(4): p399-403)
- Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmid, A., Dale, A., Hämäläinen, M.S., Marinkovic, K., Schacter, D. L. & Rosen, B.R. (2006) - Top-down Facilitation of Visual Recognition. (*Proceedings of the National Academy of Sciences of the United States of America*, vol.103 (2), p449-454)
- Bekkering, H., & Neggers, S. (2002) - Visual Search Is Modulated by Action Intentions (*Psychological Science*, vol.13 (4), p370-374)
- Block, N. (1995) - On a Confusion About a Function of Consciousness. (*Behavioral and Brain Sciences*, vol.18 (2), p227-247)
- Block, N. (2005) - Two Neural Correlates of Consciousness (*Trends in Cognitive Sciences*, vol.9(2), p46-52).
- Cohen, M. & Dennett, D. (2011) - Consciousness Cannot Be Separated from Function. (*Trends in Cognitive Sciences*, vol.15(8), p358-364)
- Dehaene, S., & Changeux, J.P. (2011) - Experimental and Theoretical Approaches to Conscious Processing (*Neuron*, vol.70(2), p200-227)
- Dehaene, S., Changeux, J.P. & Kerszberg, M. (1998) - A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks (*Proceedings of the National Academy of Sciences USA*, vol.95, p14529-14534).
- Dehaene, S., Changeux, J.P., Naccache, L., Sackur, J., & Sergent, C. (2006) - Conscious, Preconscious, and Subliminal Processing: a Testable Taxonomy (*Trends in Cognitive Sciences*, vol.10(5), p 204-211).
- Dennett, D.C. (2005) - *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. (Cambridge, MA: The MIT Press).
- Kouider, S., de Gardelle, V., Sackur, J. & Dupoux, E. (2010) - How Rich Is Consciousness? The Partial Awareness Hypothesis. (*Trends in Cognitive Sciences*, vol.14(7), p301-307).
- Lamme, V. A. F. (2004) - Separate Neural Definitions of Visual Consciousness and Visual Attention; a Case for Phenomenal Awareness (*Neural Networks*, vol.17 (5-6), p861-872).
- Sperling, G. (1960) - The Information Available in Brief Visual Presentations (*Psychological Monographs: General and Applied*, vol.74(11), p1-29)

