

Compte rendu de la conférence du professeur Stanislas Dehaene

Le 11/09/2017, **Qu'est-ce que la conscience ?**

Rédacteur : Jacques Printz – Ce CR n'engage que moi, en particulier les NB que je me suis permis d'insérer dans le résumé. Ce CR a été relu par S. Dehaene

Rappel : En plus des documents diffusés avec le bulletin N°216, voir les livres de S. Dehaene facilement accessibles ; pour ma part, j'avais lu il y a déjà un moment *Les neurones de la lecture*, et plus récemment *Le code de la conscience* dont certaines planches ont été réutilisées dans cette conférence, deux ouvrages vraiment intéressants.

En quelques mots, ce que je retiens comme points particulièrement saillants de cette conférence très intéressante. Et tout d'abord, cette interrogation initiale de SD : Peut-il y avoir une science de la conscience ?

Un débat laissé aux philosophes depuis Descartes dont SD nous rappelle la position dualiste ; c'est le débat corps/esprit, ou mind/body chez les anglo-saxons. Descartes avec sa théorie des animaux machines, inspirée des automates comme ceux que Vaucanson [1709-1782] fabriquera un peu plus tard, ne comprenait pas comment une machine pouvait « produire avec flexibilité une grande diversité de paroles et de comportements », d'où sa prise de position.

SD nous explique comment ce problème philosophique est devenu un défi expérimental, en particulier avec l'émergence d'une science informatique, et des machines informatiques beaucoup plus flexibles que tout ce qui les avaient précédées, inspirées des travaux de Turing, von Neumann, Shannon.

NB : on pourrait ajouter Wiener à la liste, avec la cybernétique. L'article provocateur de Turing, *Computing machinery and intelligence*, paru en 1950, dans la grande revue philosophique des anglo-saxons, *Mind*, traite de ce problème, et s'adresse à un public de philosophes complètement étrangers, à l'époque, à ce qu'il vient d'établir, avec Gödel ; en 1950 on compte les ordinateurs sur les doigts d'une seule main !

SD présente une série de planches montrant combien il est facile d'abuser la conscience avec les illusions visuelles dont certaines remontent aux travaux de H. von Helmholtz [1821-1894] sur la vision¹ ; la bibliothèque d'illusions s'est depuis considérablement enrichie². Ce que ces expériences montrent, c'est que le cerveau conscient est en retard sur le monde des phénomènes extérieurs. Les traitements perceptifs et les inférences correspondantes, opérés par nos organes des sens, ici la vue, sont non-conscients ; la perception consciente corrèle de façon « tardive » les entrées sensorielles, ce qui nécessite un certain temps, par exemple 300 ms.

SD montre alors que l'on peut construire des paradigmes expérimentaux qui vont permettre de manipuler la conscience en laboratoire, par exemple en jouant sur la rivalité binoculaire – nous avons tous un œil directeur, par exemple quand on vise quelque chose – et aussi avec les phénomènes de masquage temporaire qui font que certaines choses perçues de façon inconsciente ne seront finalement pas vues de façon consciente car trop rapide. Le but de ces protocoles est de découvrir de façon objective et mesurable, les signatures de la conscience en mesurant avec un appareillage ad hoc l'activité cérébrale n'importe où dans le cerveau.

¹ Pour les passionnés, mais le sujet est passionnant, voir les ouvrages d'Helmholtz, *Optique physiologique*, 2 Vol. en reprint chez Jacques Gabay ; il y a aussi un volume sur l'acoustique.

² Pour les curieux, voir <https://positivr.fr/illusions-d-optique-gif/>

SD donne l'exemple de microélectrodes qui permettent de voir/savoir si UN neurone « décharge » ou non, ce qu'il appelle un neurone unitaire ; il signale que certains neurones ne « déchargent » que lors d'une détection consciente de leur image dite « image préférée », ce qui présuppose que le neurone en question dispose dans sa structure de l'information nécessaire à cette reconnaissance.

SD explique ensuite, avec documents expérimentaux à l'appui, la mécanique à l'œuvre dans le cerveau avec les traitements de « statistiques précoces inconscientes, puis embrasement conscient » et les latences propres mesurées expérimentalement de ces mécanismes. Ce sont ces mécanismes qui expliquent les phénomènes dit de « cécité inattentionnelle » absolument spectaculaires comme ceux montrés dans le film *Whodunit* que l'on peut regarder directement sur *YouTube* [Voir <https://en.wikipedia.org/wiki/Whodunit>]. Comme il est dit dans ce film "*It is easy to miss something you are not looking for*", ce que chacun d'entre nous peut facilement expérimenter avec un peu d'introspection – une source importante de données, selon SD – via son téléphone portable !

Pour expliquer ces phénomènes, SD, en collaboration avec J-P. Changeux, a développé l'hypothèse de l'espace de travail global avec une hiérarchie perceptive du type :

Conscient → Préconscient → Subliminal.

Cet espace global, alimenté par de nombreux processeurs spécialisés qui traitent les informations non-conscientes, est localisé dans les régions préfrontale et pariétale, avec un réseau de neurones denses ayant des axones long pour distribuer ces informations globales. La conscience, nous précise SD, agit comme un goulot d'étranglement central.

SD pose ensuite la question : A quoi sert la conscience ? Un problème/question qui ne va pas de soi dit-il ! Parmi les fonctions possibles répondant à ce besoin, celle qui concerne la surveillance des opérations mentales et le repérage des erreurs est particulièrement intéressante. Comme l'explique SD, « La détection de ses propres erreurs requiert une perception consciente » ; il donne un diagramme explicatif – c'est un processus dynamique, un automate comme ceux de von Neumann – de prise de conscience de l'erreur avec un modèle de prise de décision à deux voies. L'une de ces voies utilise la représentation de l'intention d'agir sous la forme des actions motrices sous-jacentes et préalable à la décision d'agir proprement dite. La « différence » entre la voie de l'intention consciente et la voie de l'action motrice sous-jacente peut conduire à une détection d'erreur qui provoquera la prise de conscience de l'erreur.

NB : pour l'informaticien que je suis, toute cette partie évoque des mécanismes qui sont à l'œuvre dans nos systèmes informatisés depuis le début des années 1980, date à laquelle les transistors, grâce aux technologies dites VLSI, sont devenus suffisamment abondants pour compenser, via une architecture ad hoc, les effets des erreurs aléatoires de différentes natures qui peuvent se présenter, et pour améliorer la disponibilité des systèmes : conséquences, il n'y a plus de pannes visibles. Ce qui se rapproche le plus de ce qui est expliqué ici sont les processeurs de service, une machine dans la machine, dont le rôle est de surveiller en permanence ce qui se passe dans la structure organique du système à surveiller – ce qui n'est pas vu par les programmeurs d'applications externes mais que connaissent bien les programmeurs systèmes – de façon à anticiper les pannes, et prévoir les contre-mesures ; c'est ce qu'on appelle les systèmes à tolérance de pannes, dispositifs sans lesquels la disponibilité de nos machines actuelles qui intègrent des milliards de transistors et des millions de lignes de code source, et qui fonctionnent toutes en temps réel comme nos smartphones, serait catastrophique.

C'est ainsi nous dit SD que peut s'expliquer le fait expérimentalement constaté que « l'on ne voit pas ce à quoi on n'est pas attentif ».

SD explique ensuite comment le cerveau se comporte quand on veut lui faire faire deux choses en même temps, c'est-à-dire des « tâches doubles », comme la poursuite de deux cibles successives. Il met en évidence de façon expérimentale que dans une tâche double « seule l'étape tardive de la seconde tâche doit être, soit différée, soit annihilée. Ceci permet d'introduire la notion de « tâche aveugle de la conscience » qui se solde par une absence totale d'introspection concernant la cible N°2 de la seconde tâche ; celle-ci est bloquée par le processus de traitement effectué pour la 1^{ère} cible. Faire une chose à la fois pour tenir compte du goulot d'étranglement que constitue l'accès conscient à l'« espace global », sachant que cet accès est lent → ≈ 1/3 de seconde !

Les résultats expérimentaux présentés par SD montrent que l'activité cérébrale ne s'arrête jamais, même pendant une anesthésie générale. L'espace de travail global est autonome ; il n'a pas besoin d'entrées externes pour s'activer. Comme il a été indiqué, son rôle est de distribuer l'information à d'autres processeurs. Pour illustrer ce mécanisme, SD utilise la métaphore du routeur de télécommunication qui permet une « mise en série des opérations, comme le ferait une machine de Turing » précise-t-il. C'est une analogie intéressante.

NB : Là encore, pour l'informaticien que je suis, ces mécanismes évoquent beaucoup de choses dans les systèmes actuels. D'abord, plutôt que de parler de routeurs qui sont des équipements matériels « sans intelligence », c'est-à-dire sans software, quoique indispensables, mieux vaudrait parler de « *software/enterprise service bus* », ou encore de « *software pipelining* » qui correspondent assez exactement à ce que SD explique ici. Et plutôt que d'invoquer la machine de Turing qui est une abstraction d'un « humain qui calcule », aux dires de Turing lui-même, antérieure de 15 ans aux 1^{ers} ordinateurs, mieux vaudrait utiliser la métaphore des ordinateurs réels, c'est-à-dire l'architecture de von Neumann qu'on devrait d'ailleurs appeler architecture de Turing/von Neumann. Ces dispositifs système, sans lesquels Internet et le Web sont impensables, ont exactement les caractéristiques décrites par SD : ils sont autonomes, ils sérialisent, ils stockent les transactions et leurs résultats dans des caches pour accélérer les traitements répétitifs, ils réparent et « tuent » éventuellement des transactions en conflits avec celles qui précèdent pour les rejouer plus tard quand le conflit a disparu, etc.

SD conclut son exposé en évoquant des applications concrètes de sa théorie, tant au niveau clinique pour le soin des patients handicapés, qu'au niveau pédagogique pour les apprentissages chez l'enfant. Il mentionne les problèmes éthiques posés par ce qui est qualifié de « mort cérébrale », qu'il faut donc nuancer, compte tenu de ce qu'il vient d'expliquer, et les pathologies correspondantes de type coma, état végétatif, conscience minimale, etc.

Son constat final est résumé par une formule lapidaire : **on ne peut plus être dualiste !** C'est un message à destination de nos collègues philosophes et du grand public, un message par ailleurs bien compris par les physiciens, vu ce que l'on sait sur la « matière » ...

SD mentionne également, tout à la fin de son exposé, quelques travaux faits dans le cadre de l'Académie Pontificales des Sciences.

NB : J'ai eu la curiosité d'aller voir, il y a un article intéressant de SD, et d'autres également, dans *Scientific Insights into the Evolution of the Universe and of Life*, cf. <http://www.pas.va/content/accademia/en/publications/acta/evolution.html>, celui de SD est en page 394 ; il y a également dans un autre document, une réflexion sur la mort cérébrale *Why the Concept of Brain Death is Valid as a Definition of Death* dans les *Extra series* 2008 cf. <http://www.pas.va/content/accademia/en/publications/extraseries/braindeath.html>

Bref, beaucoup de matière intéressante !