



De la conscience artificielle

Catégorie : **Vie & Conscience**

Tags : **philosophie, corps, société, information**

Personnages : **Giulio Tononi, Christof Koch, Maurizio Scandella, Friedrich Nietzsche**

13 octobre 2017

Les neurosciences et les mathématiques peuvent-elles nous proposer une vision mécanisable de la conscience ? En tout cas, elles cherchent activement.

François N.

Il y a neuf ans, sur un vieux forum Yahoo, un certain François N. répondait à la question « *une machine peut-elle avoir conscience d'elle-même ?* » de la façon suivante :

N'importe quoi ! Arrête de te poser des questions pareilles. [...] Une machine ne peut avoir conscience d'elle-même car en premier lieu elle n'a pas de conscience et puis elle n'a pas de vie. C'est un robot comme un ordinateur, une automobile (assemblage de divers matériels qui, associés ensemble, constituent un système). [...] Donc une machine est simplement un système inerte et inconscient et il est impossible que cette conscience puisse surgir un jour [...] !

Il nous semble que l'opinion de François N. reste largement partagée : la conscience de soi nécessite la conscience tout court qui est un phénomène vital ne pouvant pas « *surgir de l'inerte* ». Mais c'était il y a neuf ans. Il n'est pas sûr qu'il n'ait pas *légèrement* changé d'avis depuis, surtout s'il lit les journaux et entame une co-évolution avec Google Home, Amazon Echo, ou Buddy le petit robot... Et les neuroscientifiques ne partagent plus depuis longtemps la franche conviction de François N.

Giulio T.



Giulio Tononi

En 2004, Giulio Tononi, psychiatre et neuroscientifique, connu en particulier pour ses recherches sur le sommeil, propose une théorie scientifique de la conscience, l'IIT (« *Integrated Information Theory* »), théorie toujours vigoureuse aujourd'hui. En 2016, le journal *Aeon* a interrogé Giulio Tononi et publié un article intitulé « *Consciousness creep* » (l'immixtion de la conscience) et dont voici les derniers lignes¹ :

Nous sommes déjà confrontés à des systèmes qui agissent comme s'ils étaient conscients. Notre réaction à leur égard dépend de ce que nous pensons qu'ils sont vraiment ; dès lors, des outils comme l'IIT seront nos boussoles éthiques. Tononi déclare : « La plupart des gens aujourd'hui diraient encore, **oh non, il s'agit juste d'une machine**, mais ils s'appuient sur une vision erronée de ce qu'est une machine. Ils en restent à l'image d'un truc froid, brinquebalant sur une table. Ils ne sont aucunement préparés à faire face à une machine qui les décoiffe. Quand cela arrivera – et qu'elle fera preuve d'une émotion qui vous tirera des larmes et vous fera réciter de la poésie – je pense qu'il y aura un gigantesque changement. Et les gens diront alors, **pour l'amour du ciel, comment peut-on débrancher ce truc ?** »

Turing en aurait peut-être rêvé ([Le corps de Turing](#)). Mais d'où vient à Giulio Tononi et à ses collègues une pareille conviction ?

Christof K., René D.

Disons-le d'emblée : l'IIT est une théorie axiomatique et *mathématique* de la conscience, ce qui devrait déjà éveiller les défenses naturelles de ceux qui savent ce que « théorie axiomatique » veut dire et par où ce genre de théorie se démonte.

¹ George Musser pour *Aeon* – 25 février 2016 – [Consciousness creep](#)

Cela commence ainsi : il semble impossible de nier que nous soyons conscients et en particulier conscients de nous-mêmes. Comme le dit simplement Christof Koch, un important continuateur des travaux de Tononi² :

Croire, comme certaines personnes, que la conscience est une illusion est tout à fait ridicule.

Et il convoque le fameux *cogito ergo sum* cartésien. La « démonstration » de l'existence de la conscience étant faite, celle-ci est instituée en objet scientifique légitime et peut s'offrir aux scalpels. Seul problème : cet objet reste scellé dans la boîte crânienne et ne s'est jamais manifesté en tant que tel dans les expériences d'électroencéphalographie.

La proposition de départ de Tononi est donc la suivante : plutôt que d'essayer de repérer un objet qui semble inobservable, *partons* des caractéristiques phénoménologiques supposées de la conscience (une axiomatique) et *déduisons* les principes structurels et matériels des systèmes capables de les produire. Tout système obéissant à ces principes sera dès lors considéré comme conscient.

Qu'est-ce donc qui caractérise la conscience selon Tononi ?

En premier lieu, la conscience a quelque chose à voir avec la capacité à *intégrer* de l'information. Plus précisément, la conscience résulte de la construction « à la volée » d'un état mental unique et indivisible. C'est ainsi que l'état mental (et nous ajouterions volontiers préverbal) « limace », la conscience d'une limace, émerge de l'intégration des millions de pixels de l'image rétinienne de son corps et du reflet de son mucus...

Cependant, nous disposons aujourd'hui de réseaux neuromimétiques parfaitement capables de réaliser ce type d'intégration, de « reconnaître » un visage, une empreinte, etc. Nous sommes pourtant tous d'accord : ces réseaux (qualifiés d'intelligents) ne sont que des programmes informatiques, certes sophistiqués mais démunis de toute forme de conscience. Que leur manque-t-il donc ?

Tononi propose une expérience de pensée :

Vous êtes placé devant un écran [...], et on vous demande de dire « allumé » quand il s'allume et « éteint » quand il s'éteint. Une photodiode – un composant très simple et sensible à la lumière – est également placée devant l'écran et bipe lorsque l'écran s'allume et reste silencieux lorsque l'écran s'éteint. Voici le premier problème de la conscience : quand vous faites la différence entre l'écran allumé et éteint, vous faites l'expérience consciente de « voir » la lumière ou l'obscurité. La photodiode [...] ne « voit » vraisemblablement pas consciemment la lumière et l'obscurité. Où réside donc cette différence-clé entre vous et la photodiode [...] ?

La différence essentielle réside, selon Tononi, dans le fait que la photodiode ne peut discriminer qu'entre deux situations, alors que l'expérience que l'on vous demande de vivre est l'une parmi un nombre inimaginable d'expériences que vous savez discriminer

² Antonio Regalado pour MIT Technology Review – 2 octobre 2014 – [What It Will Take for Computers to Be Conscious](#)

(par exemple, reconnaître un écran éteint, une peinture de Jackson Pollock ou un triangle rectangle). C'est ce qui explique que notre réseau de neurones « intelligent » n'est pas conscient. Bien qu'il soit capable de réaliser la prouesse de reconnaître des limaces dans toutes les configurations visuelles possibles, même les plus floues, le monde se réduit à cela : l'état « limace » (une vraie limace) ou l'état « pas limace » (un tableau de Jackson Pollock).

Une conscience émergerait ainsi d'un système capable d'intégrer et de discriminer immédiatement (dans un temps assez bref) un nombre invraisemblable d'expériences possibles. Mais en quoi consiste vraiment « l'intégration » ?

L'intégration de l'information dans l'expérience consciente est phénoménologiquement évidente : quand vous « voyez » consciemment une image, cette image est vécue comme un tout et ne peut pas être décomposée en sous-images vécues séparément. Par exemple, quel que soit l'effort que vous faites, vous ne pouvez pas faire l'expérience des couleurs d'un côté et des formes de l'autre, ou de la partie gauche du champ visuel indépendamment de la partie droite.

Même si nous pouvons *a posteriori* penser séparément la partie avant et la partie arrière de la limace, le mouvement de conscience initial les intègre comme si elles étaient causalement liées. La reconnaissance de la « limace-avant » cause et renforce la reconnaissance de la « limace-arrière » et réciproquement, tout ceci dans un même élan.

C'est donc à peu près cela, selon la théorie IIT, un système « conscient » : un système qui intègre et reconnaît, dans un élan unique, un état parmi une gigantesque quantité d'états possibles.

Il se trouve que ces principes peuvent déterminer une axiomatique mathématisée des systèmes capables de conscience³. Dès lors, par un travail mathématique classique, on déduit qu'il existe une « capacité de conscience » graduelle, que l'on peut mesurer à l'aide d'un indicateur noté Φ . Si la valeur de Φ d'un système n'est pas nulle, alors le système est « conscient ». Ainsi, si nous lisons demain dans notre journal que telle machine ou tel système est « conscient », cela peut vouloir dire que ce système obéit aux axiomes de Tononi et a une mesure $\Phi > 0$...

Nous admettons volontiers que l'IIT peut révéler les conditions *nécessaires* d'une « proto-conscience », et qu'à ce titre Φ mesure bien quelque chose d'intéressant. Mais cette théorie ne dit rien des conditions suffisantes de la « conscience », encore moins de sa fonction.

Un effleurement philosophique s'impose donc.

³ Pour une introduction scientifique et mathématique, voir Giulio Tononi et Christof Koch pour The Royal Society Publishing – 19 mai 2015 – [Consciousness: here, there and everywhere?](#)

Friedrich N., Anne B.

La conscience, et en particulier la conscience de soi, nous semble si évidente, si « déjà là », qu'on peine à la mettre à distance et à la saisir comme objet. Le *cogito ergo sum* lui-même peut paraître douteux à cause de sa réflexivité et nous sommes quelques-uns à n'avoir jamais réussi à vivre l'expérience cartésienne (je suis conscient que je suis) *sans le langage*. De ce point de vue le *cogito* est loin d'être la manifestation d'un solipsisme pur : c'est ce qui reste de l'Autre quand nous sommes seul, l'Autre avec lequel nous avons mutuellement élaboré un « moi » social (inscrit dans un corps).

En quelque sorte, la conscience c'est ce que le langage nous fait : elle n'est pas naturelle mais culturelle. Chacun peut tenter l'expérience d'une conscience sans langage, sans mots affleurant : nous approchons une forme de proto-conscience révélée par l'IIT mais guère plus.

Ce postulat n'est bien entendu pas nouveau et de nombreux philosophes l'ont établi. Dans un court essai intitulé « Nietzsche et la conscience linguistique »⁴, Maurizio Scandella, de l'université de Bologne, cite Nietzsche :

L'homme, comme toute créature vivante, pense continuellement, mais ne le sait pas ; la pensée qui devient consciente n'en est que la plus infime partie, disons : la partie la plus superficielle, la plus mauvaise, car seule cette pensée consciente advient sous forme de mots, c'est-à-dire de signes de communication, ce qui révèle la provenance de la conscience elle-même. Pour le dire d'un mot, le développement de la langue et le développement de la conscience (non pas de la raison, mais seulement de la prise de conscience de la raison) vont main dans la main.

Il manque ainsi à l'IIT l'essentiel : la saisie de la dynamique de la conscience et de sa fonction. La conscience est loin d'être seulement la saisie immédiate d'une scène, saisie que l'on peut axiomatiser.

Nietzsche allait encore plus loin comme Maurizio Scandella nous le rappelle :

La relation intersubjective est essentiellement « en dette ». La dette donne au sujet une conscience (un « je » qui se souvient de sa promesse) et un corps, c'est à dire une « garantie » de la dette.

Même notre chère Anne Barratin ([De la valeur des e-choses](#)) s'y met :

Parlez beaucoup à l'enfant de sa conscience pour qu'il sache, homme, rester sous sa domination.

Excessif ? A voir...

⁴ Maurizio Scandella– *Nietzsche et la conscience linguistique*

Machines conscientes ?

Même sans aller jusque-là, nous doutons que $\Phi > 0$ suffise à envisager un système comme conscient.

A moins que la conscience puisse devenir un jour un objet repérable, il est probable qu'un artefact ne sera « conscient » qu'à condition... que nous lui ayons assigné cette capacité ! Et sur ce point, nous sommes en quelque sorte vulnérables. A l'âge adulte nous n'en sommes plus à parler à notre doudou et à imaginer sa réponse, mais nous restons avec cette disposition apprise consistant à nous relier à toutes choses (nos semblables, des animaux, des systèmes naturels (Gaïa...), des objets, des artefacts...) en engageant notre moi conscient et en assignant à ces choses une capacité du même ordre.

La conscience est donc aussi affaire de convention : une machine consciente, c'est une machine socialement admise comme consciente, désirée, pourquoi pas, comme telle. Nietzsche n'est pas loin mais reste à distance suffisante.

Enfin, autant nous pouvons concevoir l'intérêt économique d'une machine « intelligente », autant nous notre intérêt à ce qu'elle soit « consciente » n'a rien d'évident. Ce serait même plutôt l'inverse (un animal doué de conscience, par exemple, est problématique pour l'industrie agroalimentaire). C'est pourquoi la recherche sur la conscience n'a pas pour but de la reproduire mais de l'identifier pour la soumettre.

Du point de vue du « système technicien » ([Jacques Ellul et le Système Technicien](#)), la conscience est un effet de bord *indésirable* qui advient subrepticement : « *Consciousness creep* », l'immixtion inopinée et problématique, à la Tononi, et ainsi son corollaire sensationnaliste : « *comment peut-on débrancher ce truc ?* ».