



Des zombies non-modernes

Catégorie : **Vie & Conscience**

Tags : **philosophie, robot, éthique, captologie, psychologie, algorithmes**

Personnages : **Mark Deuze, David Chalmers, Selmer Bringsjord, Luciano Floridi**

15 juin 2018

Le monde numérique produit toutes sortes de « zombies », des structures et des organisations sans conscience. En voici quelques exemples.

Le zombie hyperhybride

Les nombres, symboles et informations s'écoulent sans friction dans nos réseaux, tel un superfluide, et nous transforment¹ :

Nous passons désormais 6 heures par jour en moyenne à utiliser des appareils connectés à Internet, soit un tiers de notre temps hors sommeil ! En cumulé, plus d'un milliard d'années ont été passées en ligne en 2017.

La part prélevée sur le temps humain ([De la valeur des e-choses](#)) est phénoménale mais surtout, elle a accéléré : 7% d'internautes et 13% d'individus actifs sur les réseaux sociaux de plus en une seule année ! Plus généralement, il est devenu difficile d'envisager l'existence, même simplement matérielle, au plus bas de la pyramide de Maslow, sans appendices digitaux. L'humain nouveau est *hyperhybridé* au sens où ces appendices l'augmentent mais commencent aussi à l'ignorer². Cette hybridation débridée

¹ Flavien Chantrel pour le Blog du Modérateur – 30 janvier 2018 – [État des lieux 2018 : l'usage d'Internet, des réseaux sociaux et du mobile en France](#)

² Les Numériques – 20 janvier 2014 – [Être livré avant même d'avoir commandé ? Amazon a un brevet pour ça...](#)

pose évidemment quelques légers problèmes d'adaptation dont nous n'avons pas fini de faire le tour³ :

[...] un téléphone intelligent dans son champ de vision diminue la concentration et la capacité à échanger avec une autre personne. Même si vous n'y touchez pas ou ne le regardez pas.

Plus inquiétant, les études récentes multiplient les mises en garde concernant l'impact de ce mode hyperhybride sur la santé mentale des jeunes, notamment aux États-Unis⁴ :

Les taux de dépression et de suicide chez les adolescents ont explosé depuis 2011. Il n'est pas exagéré de décrire l'iGeneration [les enfants nés après l'an 2000] comme étant au bord de la pire crise de santé mentale depuis des décennies. L'essentiel de cette détérioration est imputable à leurs smartphones. [...] Les symptômes dépressifs ont progressé de 21% chez les garçons entre 2012 et 2015, et de 50% chez les filles... [...] En 2015, il y a trois fois plus de suicides qu'en 2007 chez les jeunes filles de 12 à 14 ans.

Mark Deuze, un spécialiste néerlandais des médias, rappelle cette évidence :

Quand nous vivons dans le media, d'une façon ou d'une autre nous devenons moins conscients de notre environnement, moins en phase avec nos sens, davantage semblables à des automates inertes [lifeless].

Des zombies ?

Le « zombie » qualifie en général une personne dont la volonté a été abolie. En l'absence de volonté, il erre – parfois de manière dangereuse – dans un environnement duquel il ne semble pas avoir conscience : le zombie *ne devrait pas être là* au sens où son « esprit » n'est pas adapté au contrôle de son corps dans cet environnement. Quelque chose de plus doit alors le maintenir dans cette existence, par exemple une batterie d'appendices techniques connectés.

L'image du zombie hyperhybride semble donc valable si on évite toute connotation fantastique. Il désigne cet être mixte dont l'état de conscience n'est plus clairement définissable et qui ne sait plus lui-même s'il se sent « bien » ou « mal ».

Le zombie de Chalmers, ou « zombie philosophique »

Présenté ainsi, il n'est pas étonnant que le zombie soit aussi une figure des sciences cognitives relative au « problème de la conscience » : qu'est-ce que la conscience ? Une machine consciente est-elle concevable ? Etc. ([De la conscience artificielle](#)).

³ Radio Canada – 10 mars 2018 – [Le malaise de la solitude à l'ère des réseaux sociaux](#)

⁴ Excellent article de Jean M. Twenge pour The Atlantic – Septembre 2017 - [Have Smartphones Destroyed a Generation?](#)

L'inconvénient de cette approche directe et logique de la conscience, c'est qu'elle finit rapidement dans d'inextricables noeuds liés à son axiome principal : je suis conscient que je suis conscient. Le problème de la conscience est pour ainsi dire assez mal né...

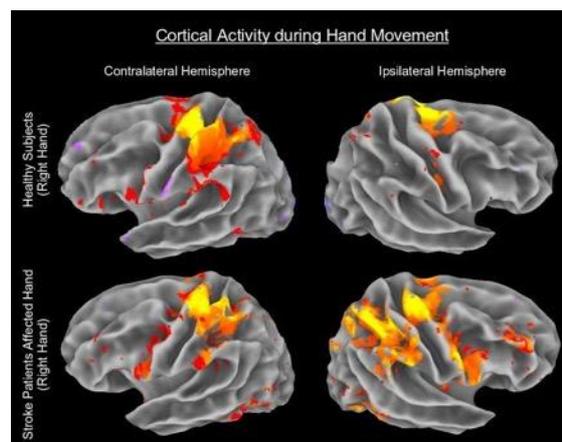
Mais les sciences cognitives, foin de prudence, tracent leur sillon... Le « zombie philosophique », repris et développé par le philosophe australien David Chalmers dans les années 1970⁵, est un pur concept logique visant à montrer que (le phénomène de) la conscience n'est pas réductible aux propriétés physico-chimiques du cerveau.



David Chalmers

Il y a pourtant toutes les raisons de croire que les neurosciences mâtinées de sciences numériques, à coups d'électromagnétocapteurs et de vérifications algorithmiques, livreront progressivement des explications causales en matière de mémoire, de rôle des émotions, d'apprentissage et, d'une façon générale, concernant tous les mécanismes physico-cognitifs de l'homme et de l'animal. C'est déjà fait, pour ainsi dire.

Mais, décidément, la conscience semble devoir résister. Comment se fait-il qu'une théorie physique (la réduction des phénomènes cognitifs aux mouvements électriques d'essaims de neurones) échoue à expliquer un phénomène qui se passe pourtant là, sous les yeux des neuroscientifiques ? Comme un mouvement de la main :



Activité corticale pendant un mouvement de la main

⁵ Wikipédia – [Zombie \(philosophie\)](#)

Chalmers raisonne de la façon suivante : rien dans notre fonctionnement ne semble devoir nécessiter la conscience. Par exemple, pour éloigner la main d'une source chaude, il n'est pas *logiquement* nécessaire d'être conscient de la chaleur. Il suffit d'un algorithme cérébral (ou ventral...) direct entre les capteurs électrochimiques de la main et les muscles du bras. Il suffirait ainsi de suffisamment d'algorithmes, innés et acquis, ou de n'importe quoi d'autre qui relie causalement nos perceptions à nos actions pour que nous agissions comme nous le faisons habituellement mais sans que nous ayons réflexivement conscience de ce que tout ceci nous fait.

De plus, et Chalmers ne le savait pas à l'époque, l'imagerie cérébrale nous apprend que la conscience (phénoménale) semble légèrement postérieure à l'action : je suis conscient que je me brûle après avoir écarté la main et non pas j'écarte la main parce que je suis conscient que je me brûle. La conscience n'est pas *causalement* reliée à l'action ; elle se manifeste « en plus ».

Il est ainsi possible de faire logiquement abstraction de la conscience et de concevoir un mode fonctionnel zombie. Le zombie est notre jumeau complet, indiscernable de nous, qui agit comme nous, mais sans conscience. Un monde zombie est donc concevable. Il est une duplication parfaite, à l'atome près, de notre réalité, mais ses occupants sont démunis de toute conscience phénoménale.

Si la conscience apparaît ainsi en surplus, de façon contingente, c'est qu'elle peut faire l'objet d'une approche scientifique propre, qu'on peut partir à la recherche de « lois psychophysiques » qui portent sur les phénomènes de la conscience en tant que tels, irréductibles à une réalité physique (et pourraient donc, au passage, s'articuler à d'autres lois de la nature que celles que nous connaissons, à d'autres structures matérielles ou virtuelles). Une nouvelle science à bâtir !

Mais ce bel édifice tient entièrement à la concevabilité du zombie. Autrement dit, Chalmers tire ses conclusions de prémisses peu stables et qui autorisent des objections elles-mêmes assez gazeuses : sans conscience, comment concevoir un être sans conscience ? Dans le monde zombie, peut-on concevoir un monde (non) zombie ? Qu'est-ce-que la causalité ? etc.

Ce raisonnement concluant que la conscience est d'une autre nature et mérite une autre science est pour le moins fragile et le principe d'une conscience artificielle basée sur une structure logico-neuronale reste sauf. C'est la position des détracteurs de Chalmers et l'intérêt de l'industrie.

Le zombie de Bringsjord



Selmer Bringsjord

Selmer Bringsjord, un cognitiviste américain, a mené en 2015 une expérience décrite dans un article au titre évocateur : « *Real Robots that Pass Human Tests of Self-Consciousness* »⁶ (à la suite des apories autour de Chalmers, un doute nous saisit⁷...).

Voici l'expérience en question.

On tape sur la tête de trois robots programmables Nao à « qui » on a précisé que cette tape peut rendre « muet » (le robot ne peut plus émettre aucun son) ou pas. On demande alors à chaque robot s'ils ont été rendus muets ou pas (« *which pill did you receive ?* ») et d'expliquer pourquoi :

[youtube https://www.youtube.com/watch?v=MceJYhVD_xY]

Le robot de droite, le seul qui n'a pas été rendu muet, commence par dire « *I don't know* » puisque a) il est programmé pour toujours répondre, comme une enceinte connectée (Google Home, Amazon Echo...) b) il ne peut pas logiquement le déduire tant qu'il n'a pas parlé. Mais s'étant « entendu » répondre, il peut en déduire qu'il n'a pas été rendu muet et corrige sa réponse en commençant par « *Sorry, I know now...* » (cette expression est complètement zombie).

Cette expérience a été suggérée par un étrange test de conscience de soi (self-consciousness) proposé par Luciano Floridi ([De l'infosphère à une éthique gazeuse](#)). Ce test est censé pouvoir permettre de distinguer zombie / non-zombie. Il consiste, dit de façon très simplifiée, en ce que de l'information sur le monde peut être tirée par un non-zombie (*self-conscious*) non seulement de son environnement mais aussi de ses propres réponses⁸.

L'article de Bringsjord commence prudemment par un démenti et une précision :

[Bringsjord] a expliqué à plusieurs reprises qu'il est impossible qu'une simple machine dispose d'une authentique conscience phénoménale, et la vraie conscience de soi [self-consciousness] exige la conscience phénoménale. Néanmoins, la structure logico-mathématique de la conscience de soi peut être établie et spécifiée, et cette spécification peut être programmée de façon à passer des tests clairs d'aptitude mentale.

Et effectivement, le robot Nao dispose d'un système de règles logiques qui lui permettent de dérouler le comportement que l'on cherche à obtenir. Nous montrons ci-dessous à quoi ressemble visuellement cette structure logico-mathématique (le « programme ») sur laquelle repose le comportement du robot « conscient de lui-même » :

⁶ Selmer Bringsjord, John Licato, Naveen Sundar Govindarajulu, Rikhiya Ghosh, Atriya Sen – Rensselaer AI & Reasoning (RAIR) Lab – [Real Robots that Pass Human Tests of Self-Consciousness](#)

⁷ Un doute tout juste rattrapé par le conditionnel... Libération, citant le magazine Quartz – 17 juillet 2015 – [Un robot pourrait avoir conscience de lui-même](#)

⁸ L. Floridi, "Consciousness, Agents and the Knowledge Game" Minds and Machines, vol. 15, no. 3-4, pp. 415-444, 2005.

Syntax

$S ::=$ Object | Agent | **Self** | Agent | ActionType | Action \sqsubseteq Event |
 Moment | Boolean | Fluent | Numeric

action : Agent \times ActionType \rightarrow Action

initially : Fluent \rightarrow Boolean

holds : Fluent \times Moment \rightarrow Boolean

happens : Event \times Moment \rightarrow Boolean

clipped : Moment \times Fluent \times Moment \rightarrow Boolean

$f ::=$ *initiates* : Event \times Fluent \times Moment \rightarrow Boolean

terminates : Event \times Fluent \times Moment \rightarrow Boolean

prior : Moment \times Moment \rightarrow Boolean

interval : Moment \times Boolean

$*$: Agent \rightarrow **Self**

payoff : Agent \times ActionType \times Moment \rightarrow Numeric

$$\frac{}{C(t, P(a, t, \phi) \rightarrow K(a, t, \phi))} [R_1]$$

$$\frac{}{C(t, K(a, t, \phi) \rightarrow B(a, t, \phi))} [R_2]$$

$$\frac{C(t, \phi) \ t \leq t_1 \dots t \leq t_n \quad K(a, t, \phi)}{K(a_1, t_1, \dots, K(a_n, t_n, \phi) \dots)} [R_3] \quad \phi$$

$$C(t, K(a, t_1, \phi_1 \rightarrow \phi_2) \rightarrow K(a, t_2, \phi_1) \rightarrow K(a, t_3, \phi_3)$$

$$C(t, B(a, t_1, \phi_1 \rightarrow \phi_2) \rightarrow B(a, t_2, \phi_1) \rightarrow B(a, t_3, \phi_3)$$

Observons nos encadrés rouges. La « conscience de soi » est la manifestation de propositions logiques manipulant un symbole... défini a priori : « Self ». Bringsjord et ses collègues ont implanté le symbole qui permet au robot de raisonner sur un « soi » et d'énoncer des phrases commençant par « je ». La conscience de soi n'émerge évidemment pas mais fait (logiquement) comme si. Cette instance de Nao est donc un zombie philosophique à l'état pur qui commence par répondre dans son langage mathématisé :

$K(I, t_4, \mathbf{happens}(\mathbf{action}(I^*, S(I^*, t_4, \mathbf{"I don't know"})), t_4))$

[je ne sais pas]

... puis se reprend (« *Sorry...* ») :

$K(R_3, t_4, \mathbf{not}(\mathbf{happens}(\mathbf{action}(R_3, \mathbf{ingestDumbPill}), t_2)))$

[je n'ai pas ingéré la pilule qui rend muet]

Il est fort probable que nous vivrons de plus en plus cernés de tels zombies numériques, à savoir d'artefacts / artifices réalisés dans l'objectif que nous puissions les désigner comme conscients par une bonne imitation à base de self. Pourquoi ? Parce que nous devons pouvoir faire confiance à des objets techniques dont le comportement est devenu intrinsèquement inexplicable (IA), ou encore interagir au niveau auquel nous pouvons les comprendre, c'est-à-dire au niveau phénoménal.

Un travail technique d'imitation cohérente et crédible est donc nécessaire pour que nous, les « utilisateurs », puissions attribuer une conscience à ces authentiques zombies et ainsi leur reconnaître la possibilité de communiquer au niveau des intentions, des choix, de la responsabilité... et le cas échéant pouvoir faire entrer ces artefacts dans le cadre de normes éthiques et juridiques.

En ce sens, le travail de Bringsjord est un bon travail d'ingénieur. Il ne prétend rien de plus d'ailleurs.

Zombies non-modernes

Ces zombies, que nous qualifierons provisoirement de « non-modernes »⁹, sont d'une espèce proliférante. Il s'agit de nos machines mais aussi de nous-mêmes et de nos organisations.



Smartphone zombies

Comme les enfants de l'iGeneration, subissant une hyperhybridation peut-être trop rapide pour une conscience de soi restée encore trop humaine. Nos ingénieurs, comme Bringsjord, semblent ici manquer de « care » (ou de conscience ?), tant l'innovation exige de saturer au plus vite l'espace économique et social.

Mais nous avons aussi d'autres quasi-victimes du nombre, comme les entreprises « zombies », qui dégagent juste de quoi rembourser les intérêts de leurs emprunts, maintenues artificiellement en vie par des taux bas, ou encore les États zombies pour les mêmes raisons. Ou encore les zombies écologiques que nous sommes devenus le 5 mai dernier en France...

La technique permet de nous maintenir provisoirement dans l'existence hors de toute logique, de toute comptabilité minimale de nos corps avec ce monde.

La réflexion éthique à laquelle tous les sachants appellent devrait donc interroger cette production industrielle de zombies, ces créations a-conscientes faisant semblant d'être conscientes.

⁹ Wikipédia – [Nous n'avons jamais été modernes](#)