

## UN PEU DE PROBABILITÉS

L'origine du bayésianisme se niche dans l'article posthume du révérend et mathématicien anglais Thomas Bayes (1702-1761), « An essay towards solving a problem in the doctrine of chances ». Cette œuvre publiée par son ami Richard Price en 1763 expose ce qui deviendra le « théorème de Bayes » :

$$P(A/B) = \frac{P(A) * P(B/A)}{P(A) * P(B/A) + P(\bar{A}) * P(B/\bar{A})}$$

Soit  $A$  et  $B$  deux événements, le théorème de Bayes permet de déterminer la probabilité de  $A$  sachant  $B$  c'est-à-dire  $P(A/B)$ , si l'on connaît les probabilités de  $A$ , de  $B$  et de  $B$  sachant  $A$ .

En d'autres termes, ce théorème décrit la méthode optimale pour mettre à jour une croyance dans des conditions d'incertitude. Il est largement utilisé en intelligence artificielle, en ingénierie comportementale, en neuroéconomie, en sciences sociales, en psychologie expérimentale, et en neurosciences cognitives pour décrire ou modéliser les comportements.

plutôt que de modifier tous ses modèles du monde. À l'inverse, lorsqu'on reçoit de nouvelles informations très précises qui contredisent ce que l'on pense, nos croyances peuvent évoluer pour s'adapter efficacement aux changements. On obtient *in fine* un système flexible s'adaptant à la structure statistique du monde au gré de nos expériences. Mais ce modèle simplifié reflète-t-il le fonctionnement réel de notre cerveau ?

## UNE HIÉRARCHIE DES CROYANCES

Ce serait trop facile ! Notre monde est trop complexe pour un algorithme aussi simple. Les signaux sensoriels que nous percevons proviennent d'un environnement dynamique en constante évolution, avec une multiplicité de structures causales imbriquées. Pour en tenir compte, la théorie du cerveau bayésien suggère que notre système nerveux traite les différents signaux sensoriels simultanément de façon « hiérarchique ». Plutôt que de comparer une seule probabilité aux preuves sensorielles, notre cerveau manipule une hiérarchie de croyances à différents niveaux d'échelle d'implication spatiotemporelle, logique et d'abstraction.

Retournons dans la jungle. La génération de la croyance qu'un jaguar est caché sous le feuillage est issue de la conjonction d'une multitude de modèles statistiques imbriqués (la probabilité qu'un animal au pelage fauve dans

la jungle soit un jaguar, la probabilité qu'un animal se cachant avant de bondir soit un jaguar, la dangerosité de l'incertitude sur la présence d'un possible jaguar...). La génération d'inférences, c'est-à-dire d'hypothèses sur le monde, à partir d'une hiérarchie de prédiction avec un large grain temporel, spatial et causal aide à réduire l'ambiguïté massive des multiples entrées sensorielles, et à obtenir une représentation plus ou moins optimale de l'environnement. De simple processus neuro-cognitif, l'inférence bayésienne devient alors un puissant mécanisme adaptatif.

## LA CLÉ DE VOÛTE DU TRAITEMENT DE L'INFORMATION

Aussi élégante soit-elle sur le plan théorique, la théorie bayésienne est-elle soutenue par l'organisation neuronale du cerveau ? Plusieurs travaux ont suggéré que les inférences seraient réalisées par un assemblage hiérarchique de connexions neuronales descendantes qui implémenteraient les prédictions, et de connexions ascendantes qui transmettraient les erreurs de prédiction. Ces prédictions et ces erreurs pourraient notamment être codées par des neurones particuliers associés à un vaste réseau de dendrites et de synapses, les cellules pyramidales.

Dans ce modèle prédictif hiérarchique, les prédictions descendantes depuis les niveaux élevés de la hiérarchie sont utilisées pour réduire les erreurs de prédiction des niveaux inférieurs. À chaque niveau, seules les erreurs de prédiction qui n'ont pas pu être expliquées par les prédictions sont transmises à l'étage supérieur, et peuvent alors être utilisées pour optimiser les croyances sur le monde.

Ces nouvelles prédictions sont ensuite transmises aux niveaux inférieurs par des connexions neuronales descendantes, pour générer de nouvelles inférences. De cette façon, toute nouvelle expérience inattendue est comparée aux connaissances antérieures à travers un vaste réseau hiérarchique associant différentes échelles temporelles, causales et spatiales. Ce processus de codage prédictif hiérarchique *médié* par la signalisation neuronale constituerait ainsi la clé de voûte de l'ensemble du traitement de l'information corticale.

Plusieurs travaux suggèrent d'ailleurs que la hiérarchie des croyances est corrélée à la complexité des représentations. Plus on monte, plus les représentations codées sont abstraites et globales, plus le niveau d'implication logique est important. Plus on descend, plus le niveau de détail se renforce, mais plus l'espace de prédiction diminue, jusqu'à arriver à des modèles prédictifs extrêmement précis. Les aires cérébrales de haut niveau dévolues aux représentations complexes envoient ainsi des signaux prédictifs aux aires sensorielles primaires, >

- déterminant en partie les motifs causaux et perceptifs attendus.

Les modèles causaux de haut niveau permettraient au cerveau de faire des prédictions sur une temporalité longue ou sur des associations causales complexes. À l'inverse, les modèles de bas niveau dans la hiérarchie prédisent, certes avec précision, la nature d'une entrée sensorielle simple, mais au mieux seulement quelques secondes à l'avance. Le cerveau ne peut déduire les conséquences de ses actions que s'il peut modéliser l'avenir. Cette capacité de faire des prédictions sur des échelles temporelles longues est ainsi primordiale pour définir les actions à réaliser. La structure hiérarchique lui permet de faire des prédictions sur des choses qui ne se sont jamais produites et pourraient ne jamais se produire : une forme de simulation d'un monde contrefactuel.

Pour le philosophe Jakob Hohwy, le cerveau serait ainsi un «miroir de la nature», simulant à chaque instant la structure causale du monde en fonction des prédictions de ses caractéristiques statistiques. Le codage prédictif et la théorie du cerveau bayésien pourraient ainsi offrir une conception globale de notre fonctionnement cérébral, avec une portée explicative aussi importante que celle de la psychanalyse au siècle précédent.

## D'UNE CROYANCE À L'AUTRE

L'un des aspects les plus importants du réseau prédictif est la notion d'implication logique. Elle désigne le phénomène par lequel une croyance donnée peut en impliquer un ensemble d'autres. On peut l'illustrer avec la notion de transitivité du philosophe néerlandais Bas Van Fraassen : si je crois qu'un vase se casse lorsqu'il tombe sur le sol, alors je crois aussi qu'il se casse quand on le projette contre le mur, lorsqu'on le frappe avec un marteau... Dans ce modèle prédictif, la hiérarchie de représentations définit des relations complexes d'implication logique qui assurent une cohérence globale au sein de notre réseau de croyances. Ainsi, si je crois qu'un jaguar a quatre pattes, je crois aussi qu'il n'a pas cinq pattes, pas six pattes, et une infinité d'autres croyances associées. Même si je ne suis pas conscient de ces croyances implicites, elles existent du fait de la structure logique de mon réseau de croyances, constituant une forme d'inconscient «bayésien».

Ce processus est toutefois complexe à modéliser, car à propos d'une seule expérience ou d'une seule croyance, on peut souvent générer une infinité de relations d'implication logique. Notre cerveau doit, à chaque instant, faire le tri entre toutes les hypothèses envisageables à partir des expériences passées. Par exemple, si je crois que tous les objets en cuivre conduisent

l'électricité, je peux rationnellement croire que tous les futurs objets en cuivre que je rencontrerai conduiront l'électricité. En revanche, si je crois que tous les objets posés sur mon bureau ce matin peuvent conduire l'électricité, je ne dois pas pour autant me mettre à croire que tous les objets qui seront sur mon bureau au cours des prochains mois seront conducteurs ou que tout ce qui est conducteur se trouve sur mon bureau.

Notre cerveau doit ainsi représenter précisément le type d'implication logique entretenu par chacun des modèles internes qu'il entretient au sein de sa hiérarchie de croyances. Cette distinction est d'autant plus importante que la plupart de ces croyances ne sont pas conscientes et pourtant interagissent activement avec les perceptions, les pensées et les actions. La hiérarchie de croyances que nous avons détaillée représente ainsi une «matrice» déterminant la cognition, un modèle qui rappelle étrangement le concept d'inconscient mobilisé par la psychanalyse freudienne.

## DE FREUD À BAYES

Lorsqu'il définit son triptyque conscient-inconscient-préconscient, Sigmund Freud s'inscrit dans le sillage d'illustres prédécesseurs (Thomas Laycock, Wilhelm Wundt...). Il postule que des processus cognitifs inconscients peuvent déterminer causalement des processus conscients et que des processus conscients peuvent devenir inconscients, c'est-à-dire que tout ce qui ne relève pas de la conscience à un moment donné reste silencieusement actif et peut continuer à interférer avec les processus conscients.

L'influence causale de l'inconscient sur le conscient a depuis été illustrée par de nombreuses preuves expérimentales (*voir Empêcher l'inconscient de dicter sa loi*, par J. Bargh, page 52). Cependant, une multitude de facteurs épistémologiques ont provoqué la scission entre les théories neurocognitives de l'inconscient et la théorie freudienne, favorisant l'émergence de concepts alternatifs comme celui d'«inconscient cognitif», dont la formulation bayésienne est l'une des nuances.

Malgré cette dissonance, un certain nombre de points d'achoppement ont été soulignés entre l'inconscient freudien et cognitif. Dans l'hypothèse du cerveau bayésien hiérarchique, l'inconscient est «structurel» au sens où il repose sur l'architecture fonctionnelle du cerveau, c'est-à-dire l'organisation et la connectivité des neurones et des synapses. Les représentations qui y sont codées n'ont pas vocation à devenir directement conscientes, mais elles participent néanmoins à la conscience à différents niveaux de la hiérarchie corticale. Elles alimentent à chaque instant la conscience d'une multitude de