

Le cerveau est-il une machine Bayésienne?

Peggy Seriès

Département d'Informatique, Université d'Edimbourg, Edimbourg, Royaume-Uni.

Correspondance:

Dr Peggy Seriès,
Institute for Adaptive and Neural Systems,
10 Crichton Street,
University of Edinburgh,
Edinburgh, EH8 9AB, UK.
pseries@inf.ed.ac.uk

Résumé

Une idée de plus en plus populaire en sciences cognitives et neurosciences computationnelles est que le cerveau pourrait être comparable à une machine probabiliste ou « Bayésienne ». Cette approche est devenue omniprésente dans l'étude du comportement mais aussi de plus en plus influente dans l'étude du substrat neuronal associé. Nous expliquons ce que signifie précisément cette hypothèse pour les sciences cognitives et décrivons son utilisation pour l'étude de la perception et la prise de décision. Nous discutons ses implications et ses limites et l'implémentation possible de processus Bayésiens dans le cerveau.

Abstract

A growing theory in cognitive science and computational neuroscience is that cognition can be successfully described using probabilistic or "Bayesian models". This approach has become mainstream in experimental psychology and started to influence the study of the neural substrate as well. We describe this hypothesis and review studies, from our laboratories and others, using this approach in human perception and decision-making. We discuss implications and limits of the "Bayesian Brain hypothesis" and address how Bayesian inference processes could be implemented in the brain.

Notice Biographique

Après une formation d'ingénieur et en intelligence artificielle (ISEP 1996; MSc Université de Sussex), P. Seriès a effectué sa thèse de Biomathématiques (Paris VI) à l'UNIC (CNRS, Y. Frégnac) et sa recherche postdoctorale à l'Université de Rochester (A. Pouget), UCL (Gatsby) et NYU (E. Simoncelli). Depuis 2007, P. Seriès est enseignante-chercheuse à l'Université d'Edimbourg. Elle s'intéresse au développement de modèles Bayésiens pour la perception, la prise de décision et la psychiatrie.

1. Introduction

Les neurosciences théoriques et computationnelles utilisent des outils de modélisation mathématique et de simulation informatique pour étudier le fonctionnement du cerveau et les comportements qu'il engendre. Un développement récent de ce domaine concerne l'utilisation de modèles probabilistes et l'idée selon laquelle le cerveau pourrait être comparable à une machine Bayésienne.

Les origines de cette idée sont quelquefois attribuées à Hermann von Helmholtz¹. Etudiant l'œil humain et observant combien cet instrument optique était imparfait, von Helmholtz a proposé que la perception visuelle résultait d'un processus d'induction inconsciente ayant lieu dans le cerveau. Ce processus permettrait d'interpréter les entrées visuelles, généralement incomplètes et ambiguës, en utilisant des connaissances antérieures. L'approche Bayésienne, empruntant des outils provenant de l'apprentissage automatique, fournit un cadre théorique adapté pour décrire cette hypothèse de manière précise et a donné lieu à des modèles mathématiques et computationnels influents pour la psychologie expérimentale et les neurosciences. L'idée est que le cerveau manipulerait constamment des hypothèses et effectuerait des calculs impliquant le degré de confiance qui doivent leur être accordées. En termes mathématiques, dire qu'un système effectue une inférence Bayésienne signifie qu'il met à jour la probabilité $P(\text{hypothèse} \mid \text{données})$ qu'une hypothèse soit vraie sachant que certaines données ont été mesurées, en exécutant la règle de Bayes:

$$P(\text{hypothèse} \mid \text{données}) = \frac{P(\text{données} \mid \text{hypothèse})P(\text{hypothèse})}{P(\text{données})} \quad (1)$$

Pour le système visuel, par exemple, l'hypothèse peut correspondre à la présence d'une cible visuelle (par ex : « il y a un ours à ma poursuite » -- tâche de détection), ou à la valeur d'un stimulus donné (« la vitesse à laquelle cet ours se déplace est de 30 km/h » -- tâche d'estimation) et les données correspondent aux informations fournies par les entrées visuelles. $P(\text{données} \mid \text{hypothèse})$ mesure la compatibilité des informations recueillies avec l'hypothèse et est appelée la fonction de vraisemblance. La distribution de probabilité $P(\text{hypothèse})$ est appelée la distribution de probabilité *a priori*. Elle correspond aux attentes *a priori* concernant la validité de l'hypothèse, formées avant de recevoir les entrées visuelles et fondées sur des expériences antérieures (par ex. la probabilité d'observer un ours dans la campagne écossaise). La probabilité *a priori* joue un rôle important pour résoudre l'ambiguïté présente dans les données en situations d'incertitude. $P(\text{données})$ joue le rôle d'un facteur de normalisation, pour que la probabilité $P(\text{hypothèse} \mid \text{données})$ soit comprise entre 0 et 1.

Certains auteurs² distinguent deux hypothèses distinctes pour ce qui est de l'application de ce théorème aux neurosciences. La première est "l'hypothèse du cerveau Bayésien"³: l'idée selon laquelle le théorème de Bayes fournit un cadre théorique adapté pour décrire et comprendre les performances humaines, sur le plan comportemental. La deuxième est "l'hypothèse du codage Bayésien"⁴: l'idée selon laquelle l'activité neuronale représenterait l'information de manière probabiliste et effectuerait des calculs utilisant des densités de probabilité ou des approximations de densités de probabilité.

Nous résumons ici un ensemble de travaux qui illustrent ces hypothèses⁵. La première partie concerne plus particulièrement "l'hypothèse du cerveau Bayésien" et son application à la perception sensorielle où cette approche est devenue omniprésente. Nous décrivons ensuite l'utilisation de cette approche dans le domaine de la prise de décision et son potentiel prometteur pour la psychiatrie. Nous

¹ H. von Helmholtz, "Concerning the perceptions in general". Treatise on physiological optics, vol. III, 3rd edn, 1866 (traduction vers l'anglais de J. P. C. Southall, Opt. Soc. Am. Section 26, 1925); voir aussi: G. Westheimer, "Was Helmholtz a Bayesian? A Review". Perception, 37, 2008.

² Par exemple, D. Knill & A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation". Trends Neurosci, 27(12): 2004, 712–719.

³ en anglais: "the Bayesian Brain Hypothesis".

⁴ en anglais: "the Bayesian Coding Hypothesis".

⁵ Pour des revues en anglais, voir aussi I. Vilares & K. Kording, "Bayesian models: the structure of the world, uncertainty, behavior, and the brain". Ann N Y Acad Sci, 1224, 2011, 22–39; M. Colombo & P. Seriès, "Bayes in the brain – on Bayesian modelling in neuroscience". British Journal for the Philosophy of Science, 63, 2012, 697–723; A. Pouget, J.M. Beck, W.J. Ma & P.E. Latham, "Probabilistic brains: knowns and unknowns." Nat. Neuroscience, 16(9), 2013, 1170-8 ainsi que le cycle de conférences donné par Stanislas Dehaene au Collège de France <http://www.college-de-france.fr/site/stanislas-dehaene/course-2012-01-10-09h30.htm#%7Cm=course%7Cq=/site/stanislas-dehaene/course-2011-2012.htm>

discutons enfin "l'hypothèse du codage Bayésien", les défis épistémologiques associés et les critiques rencontrées.

2. Bayes et la psychophysique sensorielle.

2.1 L'intégration sensorielle comme " preuve " du calcul Bayésien.

La manière dont le cerveau intègre différentes sources d'information est souvent considérée comme une preuve que le cerveau peut effectuer des calculs Bayésiens.

Pour comprendre ce résultat, considérons par exemple une situation où l'on cherche à déterminer la position d'un prédateur potentiel et deux types d'information sont disponibles à un instant donné, l'une visuelle (V), par exemple le mouvement d'un feuillage, et l'autre auditive (A), par exemple un grognement ou cri émis par l'animal (Figure 1A, B). Comment notre cerveau intègre-t-il ces deux signaux pour estimer la position X de cet animal?

Le calcul Bayésien nous informe sur la manière dont les informations doivent être combinées pour que l'estimation $P(X|V,A)$ soit optimale. L'idée est que les deux sources d'informations doivent être combinées en les pondérant par leurs fiabilités relatives. Par exemple, si l'information visuelle est très diffuse et l'information auditive plus fiable, l'information auditive doit avoir plus d'influence sur le résultat final (Figure 1C).

< Figure 1 environ ici >

La démonstration de ce résultat est très simple. En effet, la formule de Bayes (équation 1) nous indique que l'on peut représenter la distribution de probabilité *a posteriori* $P(X|V,A)$ comme le produit de la fonction de vraisemblance avec la distribution *a priori*:

$$P(X|V,A)=P(V,A|X)P(X)/ \text{normalisation} \quad (2)$$

La fonction de vraisemblance $P(V,A|X)$ indique la probabilité des mesures V et A recueillies si nous savions que le prédateur était à une certaine localisation X. Nous ignorons la distribution *a priori* $P(X)$ pour le moment -- nous y reviendrons dans la section suivante -- en considérant que toutes les positions sont équiprobables *a priori*. Une hypothèse raisonnable est de considérer que le bruit des deux systèmes sensoriels est indépendant. La fonction de vraisemblance $P(V,A|X)$ peut alors être factorisée:

$$P(V,A|X)=P(V|X)P(A|X) \quad (3)$$

Un système Bayésien combine donc les probabilités conditionnelles des deux mesures en les multipliant. Si les deux fonctions de vraisemblance sont Gaussiennes, centrées sur les mesures V et A et écart-types s_v et s_a , on peut facilement montrer que la distribution de probabilité *a posteriori* $P(X|V,A)$ est aussi Gaussienne avec une moyenne (ou un maximum) m_p donnée par:

$$m_p^2 = \frac{s_a^2}{s_v^2 + s_a^2} V + \frac{s_v^2}{s_v^2 + s_a^2} A \quad (4)$$

et un écart-type s_p égale à :

$$s_p^2 = \frac{s_v^2 s_a^2}{s_v^2 + s_a^2} \quad (5)$$

Ces prédictions indiquent que si l'estimation effectuée par le cerveau est donnée par le maximum de la distribution *a posteriori*⁶ obtenue par la règle de Bayes, elle doit correspondre à une moyenne des

⁶ Celle-ci correspond aussi au maximum de vraisemblance dans ce cas, où l'on considère que la distribution *a priori* est uniforme.

estimations unimodales pondérées par leurs fiabilités respectives (Eq. 4). De plus, sa précision doit être meilleure que celle obtenue pour chacune des modalités en isolation (Eq. 5).

Ces prédictions sont assez facilement testables expérimentalement. En effet, en présentant une seule de ces modalités sensorielles à la fois (par exemple l'information visuelle puis l'information auditive), on peut mesurer les estimations et les précisions S_A et S_V avec lesquelles chacune de ces informations est perçue. On peut alors présenter les deux modalités simultanément, mesurer à nouveau l'estimation et la précision de l'estimation et comparer le résultat à celui prédit par les équations (3) et (4).

C'est exactement ce qu'ont testé Ernst et Banks (2002)⁷ en utilisant les modalités visuelles et tactiles. La tâche des sujets consistait à estimer la hauteur d'une arête en relief, qu'ils pouvaient voir (de manière plus ou moins précise, Figure 1C) et toucher grâce à un système de réalité virtuelle. Ernst et Banks montrent que la précision mesurée lorsque les deux modalités sont présentes suit précisément la prédiction donnée par les équations (3) et (4). Ils en déduisent que les humains combinent les informations visuelles et tactiles de manière "statistiquement optimale" et "similaire à un estimateur du maximum de vraisemblance".

Ce résultat a ensuite été reproduit par de nombreuses études et différentes modalités⁸. Ce phénomène explique aussi l'effet McGurk (une situation expérimentale où l'on présente une vidéo d'une personne prononçant un phonème (par exemple /ga/), alors que la bande son diffuse un autre phonème (par exemple /ba/) -- on perçoit alors un phonème intermédiaire (/da/)). Il peut aussi expliquer le cas extrême du ventriloquisme, où la source de la voix est attribuée à la bouche de la marionnette, qui produit une information visuelle très fiable, alors que l'information auditive produite par le ventriloque est très diffuse.

Ces résultats ont été très influents, et il est souvent considéré -- comme l'écrivent Knill et Pouget⁹ -- que l'intégration des signaux sensoriels fournit "la preuve la plus convaincante de l'hypothèse du codage Bayésien". Ils marquèrent le début de l'engouement pour les modèles Bayésiens en sciences cognitives.

2.2 Quels sont les probabilités de distributions *a priori* du système perceptif?

L'approche Bayésienne prédit non seulement la manière dont différents types de signaux simultanés doivent être combinés, mais aussi comment inclure toute connaissance ou opinion qui serait antérieure aux mesures. Ces informations sont représentées sous la forme de probabilités *a priori* (Equation 1). Récemment, un axe de recherche s'est intéressé plus directement à cette notion dans l'étude de la perception: si le cerveau utilise des distributions *a priori*, quelles sont ces distributions et quel est leur rôle dans la perception?

Selon le théorème de Bayes, ce sont dans les situations de forte ambiguïté ou d'incertitude que les distributions *a priori* sont susceptibles d'influencer les processus d'estimation. Dans le domaine de la perception visuelle, de nombreuses études ont montré que les illusions visuelles peuvent être utilisées pour révéler les attentes *a priori* ou hypothèses implicitement faites par le cerveau lorsque plusieurs interprétations de l'environnement sont possibles. Un *a priori* perceptif classiquement cité dans ce contexte concerne l'attente que la lumière vient "d'en haut". Bien que certains aspects de cet *a priori* restent débattus, il est reconnu qu'il joue un rôle déterminant dans l'interprétation des formes et dans les performances de recherche visuelle pour les objets qui projettent une ombre. Par exemple, dans la Figure 2A, le lecteur percevra un seul objet concave (un "creux") au milieu d'objets convexes (des "bosses"). Cette interprétation est liée au fait que vous faites l'hypothèse implicite que la lumière vient d'en haut de l'image. Mais retournez la page et votre perception est inversée: vous percevez maintenant une "bosse" au milieu de "creux".

De manière similaire, les humains semblent avoir une préférence *a priori* pour la symétrie¹⁰ (Figure 2B), la continuité dans l'espace et dans le temps¹¹, et pour les orientations cardinales¹². Nous nous attendons

⁷ M.O. Ernst & M.S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion". *Nature*, 415(6870), 2002, 429-33.

⁸ Pour une revue, voir par exemple: D.E. Angelaki, Y. Gu & G.C. DeAngelis, "Multisensory integration: psychophysics, neurophysiology and computation". *Curr Opin Neurobiol*, 19(4), 2009, 452-8.

⁹ D. Knill & A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation". *Trends Neurosci*, 27(12): 2004, 712-719.

¹⁰ D.C. Knill, "Learning Bayesian priors for depth perception". *J Vis*, 7(8), 2007, 13.

¹¹ O. Schwartz, A. Hsu & P. Dayan, P. "Space and time in visual context". *Nat Rev Neurosci*, 8(7): 2007, 522-535.

¹² A.R. Girshick, M.S. Landy & E.P. Simoncelli, « Cardinal rules: visual orientation perception reflects knowledge of environmental statistics ». *Nat Neurosci*, 14(7), 2011, 926-932.

aussi à ce que les objets soient convexes et que les images de fond soient colorées de manière homogène¹³, que les objets se déplacent lentement, ou restent statiques, que le regard des autres soient dirigé vers nous¹⁴, et que les visages soient des surfaces convexes (voir l'illusion dit « du masque creux »¹⁵).

Ces *a priori* sont utilisés par le cerveau pour interpréter l'image visuelle et peuvent être justifiés de manière théorique par le fait que la plupart des objets sont conformes à ces attentes. Ces *a priori* ont ainsi tous été modélisés en utilisant l'approche Bayésienne. Celle-ci montre qu'en moyenne, l'utilisation de ces *a priori* permet au système perceptif d'interpréter les entrées visuelles de manière optimale. Cependant, en situation de forte ambiguïté, lorsque les objets ne sont pas conformes aux statistiques moyennes, l'utilisation de ces *a priori* peut engendrer des illusions perceptives: nous percevons alors la réalité comme étant plus similaire à celle que nous attendons *a priori* que ce qu'elle est vraiment. Les objets paraissent plus lents, plus symétriques, plus réguliers dans l'espace et le temps etc.

< Figure 2 environ ici >

Un *a priori* qui a été particulièrement étudié et auquel nous nous sommes particulièrement intéressé est l'attente *a priori* que les objets soient statiques ou se déplacent lentement (un *a priori* « pour les vitesses lentes »). L'existence de cet *a priori* a d'abord été introduite comme une hypothèse élégante pour expliquer de manière unifiée un grand nombre d'illusions concernant la direction d'objets en mouvement; par exemple, le fait qu'une ligne en mouvement derrière une ouverture est toujours perçue comme se déplaçant perpendiculairement à son orientation (le « problème de l'ouverture¹⁶) ou le fait que la direction du mouvement d'un losange semble dépendre de son niveau de contraste. Weiss et coll. (2002)¹⁷ ont proposé un modèle Bayésien du mouvement visuel qui suppose que les mesures locales des images sont bruitées et que les mouvements lents sont *a priori* plus plausibles que les mouvements plus rapides (un *a priori* Gaussien centré sur une vitesse de 0°/s) -- une hypothèse raisonnable dans un monde où la plupart des objets sont immobiles. Ils ont montré que ce modèle pouvait rendre compte d'un grand nombre d'illusions observées en psychophysique, qui jusque-là n'avaient été expliquées que par des modèles différents et spécifiques à chaque illusion. Ces auteurs proposèrent donc que les illusions de mouvement visuel ne soient « pas le résultat de calculs bâclés par certains composants du système visuel, mais plutôt le résultat d'une stratégie computationnelle cohérente qui est optimale sous des hypothèses raisonnables ». Ils concluent finalement que les illusions visuelles doivent être considérées comme des « perceptions optimales ».

Mesure chez le sujet individuel.

D'importantes questions subsistent: d'où viennent ces *a priori*? Sont-ils différents pour différents individus? Et si c'est le cas, suffisent-ils à expliquer les différences individuelles de manière quantitative? Peut-on utiliser les performances des participants pour mesurer leurs croyances *a priori* ?

Weiss et coll. (2002) avaient construit leur modèle en utilisant un *a priori* dont la forme était choisie de manière arbitraire (gaussienne, par simplicité), mais dont ils ont montré qu'il pouvait expliquer les performances visuelles moyennes, de manière qualitative. Récemment, un certain nombre de laboratoires ont montré que les distributions *a priori* pouvaient être mesurées pour chaque individu quantitativement. La méthodologie générale consiste à partir de l'hypothèse que les performances de chaque individu sont le résultat d'une inférence Bayésienne, où l'on définit (de manière plus ou moins rigide) le modèle de bruit pour le processus d'estimation sensorielle et de la réponse motrice, les probabilités de distributions *a priori* et la fonction de coût. Par exemple, la distribution *a priori* peut être définie comme un mélange gaussien dont les paramètres sont libres. Les paramètres de ce modèle sont ensuite ajustés pour chaque sujet pour rendre compte de ses performances individuelles.

¹³ D. Goldreich & M.A. Peterson, "A Bayesian observer replicates convexity context effects in figure-ground perception". *Seeing and Perceiving* 25(3-4), 2012, 365–395.

¹⁴ I. Mareschal, A.J. Calder & C.W. Clifford, "Humans have an expectation that gaze is directed toward them". *Curr Biol.* 23(8), 2013, 717-21.

¹⁵ R. Gregory, *The Intelligent Eye*. London: Weidenfeld and Nicolson. 1970.

¹⁶ http://upload.wikimedia.org/wikipedia/commons/f/f0/Aperture_problem_animated.gif

¹⁷ Y. Weiss, E.P. Simoncelli & E.H. Adelson. "Motion illusions as optimal percepts". *Nat Neurosci*, 5(6), 2002, 598–604.

En utilisant un modèle plus flexible que le modèle gaussien pour la distribution *a priori*, Stocker et Simoncelli (2006)¹⁸ ont pu ainsi montrer qu'au niveau individuel, l'*a priori* pour les vitesses lentes n'est pas gaussien mais plutôt caractérisé par une distribution de loi de puissance avec une forte variabilité interindividuelle. À la suite de cette étude, un certain nombre de laboratoires (dont le nôtre) cherchent à comprendre si ces distributions *a priori* individuelles peuvent expliquer la variabilité observée pour d'autres tâches ou stimuli que ceux qui sont utilisés pour les mesurer. Par exemple : utilise-t-on le même *a priori* pour estimer la vitesse d'objets très différents, par exemple le déplacement des nuages et celui des cyclistes, ou existent-t-il des *a priori* spécifiques à chaque objet ? Pour certains chercheurs, cette propriété de généralisation des *a priori* perceptifs serait une preuve convaincante de la validité de l'approche Bayésienne¹⁹.

Le fait que les distributions *a priori* puissent être extraites à partir des performances comportementales est un développement important. En effet, il suggère que les modèles Bayésiens peuvent être utilisés comme méthode de rétro-ingénierie pour découvrir les modèles internes et stratégies cognitives utilisés par différents individus.

Distributions *a priori* et statistiques naturelles.

Une question qui se pose naturellement est de savoir si les distributions *a priori* mesurées chez les individus sont conformes aux statistiques de l'environnement, une condition généralement considérée nécessaire pour que l'inférence soit optimale²⁰. Par exemple, nos attentes *a priori* pour les vitesses lentes reflètent-elles précisément la distribution des vitesses des objets rencontrés dans le passé dans notre environnement ?

Il est difficile de répondre à cette question pour l'*a priori* pour les vitesses lentes. En effet, il est difficile de mesurer la distribution statistique des vitesses perçues par la rétine car celles-ci dépendent non seulement des statistiques des images naturelles, mais aussi de l'effet relatif des mouvements du corps, de la tête et des yeux. Cependant, Girshick et al (2012)²¹ ont pu répondre à cette question dans le domaine des orientations visuelles. En mesurant les performances de sujets qui devaient comparer les orientations de différentes lignes visuelles, ces auteurs ont trouvé que les participants présentaient de forts biais vers les orientations cardinales lorsque les stimuli étaient incertains, suggérant une distribution *a priori* favorisant ces orientations. Ils ont ensuite mesuré la distribution des orientations locales dans un grand nombre de photos d'images naturelles et ont trouvé que celle-ci était fortement non-uniforme, avec une dominance d'orientations cardinales. Ils ont finalement montré que les distributions *a priori* mesurées à partir des performances visuelles des participants étaient conformes aux distributions mesurées dans les images naturelles.

Les données recueillies dans cette étude parmi d'autres²² sont compatibles avec l'idée que les distributions *a priori* utilisées par le cerveau sont une approximation, plausible mais imparfaite, des distributions naturelles. Mais comment ces distributions *a priori* sont-elles apprises ? Et avec quelle constante de temps ? Sont-elles rigidement implémentées dans le cerveau, acquises à travers le code génétique, assimilées pendant le développement, ou bien leur apprentissage est-il un processus continu, actif à tout moment de notre existence ?

Les travaux existant tendent à montrer que l'apprentissage des statistiques de l'environnement a lieu de manière continue et inconsciente, même pour les caractéristiques de l'environnement les plus simples. Par exemple, nous avons mené une étude où l'on demandait à des participants d'observer un nuage de points en mouvement cohérent et d'effectuer deux tâches visuelles consécutives : ils devaient d'abord indiquer la direction du mouvement à l'aide d'un pointeur (estimation), puis indiquer s'ils avaient réellement perçu les points (le contraste est souvent très faible et les points sont absents durant certains essais). Cependant, ce que les sujets ne savaient pas, c'est que deux directions étaient beaucoup plus fréquentes que les autres (c.-à-d.

¹⁸ A.A. Stocker & E.P. Simoncelli, « Noise characteristics and prior expectations in human visual speed perception ». *Nat Neurosci*, 9(4), 2006, 578–585.

¹⁹ L.T. Maloney & P. Mamassian, « Bayesian decision theory as a model of human visual perception: testing Bayesian transfer ». *Vis Neurosci*, 26(1), 2009, 147–155.

²⁰ W.J. Ma, « Organizing probabilistic models of perception ». *Trends Cogn Sci*, 16(10), 2012, 511-518; mais voir aussi J. Feldman, « Tuning Your Priors to the World », *Topics Cogn. Sci*, 5(1), 2013, 13-34.

²¹ A.R. Girshick, M.S. Landy & E.P. Simoncelli, « Cardinal rules: visual orientation perception reflects knowledge of environmental statistics ». *Nat Neurosci*, 14(7), 2011, 926–932.

²² Voir par exemple: R. Zhang, O. S. Kwon & D. Tadin, « Illusory movement of stationary stimuli in visual periphery: evidence for a strong centrifugal prior ». *J Neurosci*, 33, 2013, 4415-4423.

toutes les directions de mouvement n'étaient pas équiprobables). Bien que cette connaissance reste inconsciente, nos expériences montrent que les sujets apprennent ces statistiques en quelques minutes et que cet apprentissage influence fortement leur perception: ils deviennent plus rapides et plus efficaces à détecter les directions les plus fréquentes, mais tendent aussi à percevoir les autres directions comme étant plus proches des directions les plus fréquemment présentées que ce qu'elles sont vraiment. De plus, lorsqu'il n'y a rien à l'écran, les sujets ont tendance à "halluciner" les directions les plus fréquentes²³. Finalement, nous avons pu montrer que ces performances étaient décrites de manière satisfaisante par un modèle Bayésien dont la distribution *a priori* est une approximation des statistiques de l'environnement, mesurable chez chaque sujet individuel. Ces travaux suggèrent ainsi que le cerveau construit en permanence de nouvelles distributions *a priori* pour les caractéristiques de l'environnement dans lequel il est placé. D'autres études ont aussi montré que les *a priori* qui semblent correspondre aux statistiques du monde naturel, comme l'attente que la lumière vient d'en haut²⁴, ou que les objets se déplacent lentement²⁵, peuvent aussi être révisés de manière dynamique en fonction des statistiques du contexte.

2.3 Prise de décision et psychiatrie computationnelle.

Les modèles Bayésiens sont aussi devenus omniprésents dans l'étude de la prise de décision, où ils fournissent un point de comparaison précis pour l'évaluation des performances humaines ou animales, sains ou malades.

Un exemple particulièrement important est la modélisation des tâches de choix forcé alternatif. Par exemple, une tâche souvent utilisée en neurosciences expérimentales consiste à présenter un nuage de points aléatoires en mouvement sur un écran. L'on demande au sujet d'indiquer si le nuage se déplace vers la gauche ou vers la droite (Figure 3A). Dans cette situation, la stratégie de décision Bayésienne peut être décrite quantitativement comme en processus en deux étapes successives : dans un premier temps, une étape d'intégration de l'évidence sensorielle D au cours du temps où sont calculées les probabilités que chaque hypothèse (H_1 : vers la droite, ou H_2 : vers la gauche) soit correcte. Dans un deuxième temps, ces hypothèses sont comparées pour donner lieu à la décision ($P(H_1|D)$ vs $P(H_2|D)$). De manière équivalente, le logarithme du rapport de ces probabilités peut-être comparé à un seuil de décision qui inclue les probabilités *a priori* et les coûts associés aux erreurs éventuelles dans la décision. Il a été montré que dans certains contextes, ce modèle Bayésien optimal (connu aussi sous le nom de "test séquentiel du rapport de vraisemblance") est équivalent à certains modèles proposés auparavant en psychologie pour décrire les performances humaines, appelés modèles de diffusion²⁶. Dans ces modèles, la décision est représentée par la trajectoire d'une particule qui se déplace entre deux limites représentant les deux choix (Figure 3B). En partant d'un point donné, la particule accumule des informations en faveur des différentes hypothèses, et chemine petit à petit vers l'une ou l'autre des limites. La décision est prise lorsqu'une des limites est atteinte. L'approche Bayésienne est non seulement compatible avec ce modèle mais permet de le contraindre et de l'interpréter de manière normative²⁷. Ce modèle est d'autant plus séduisant que de nombreuses études neurophysiologiques ont montré que les dynamiques des activités neuronales dans des aires corticales comme le cortex intrapariétal latéral (LIP) pourraient implémenter ce processus de manière relativement explicite²⁸. Ces résultats suggèrent qu'au moins dans des situations simples, la prise de décision est décrite de manière satisfaisante par un modèle Bayésien.

²³ M. Chalk, A. R. Seitz & P. Seriès, "Rapidly learned stimulus expectations alter perception of motion". *J Vis*, 2010, 10(8):2. Voir aussi : N. Gekas, M. Chalk, A.R. Seitz & P. Seriès, "Complexity and specificity of experimentally induced expectations in motion perception". *Journal of Vision*, 13(4), 2013.

²⁴ W.J. Adams, E.W. Graf & M.O. Ernst, "Experience can change the 'light-from-above' prior". *Nat Neurosci*, 7(10), 2004, 1057– 1058.

²⁵ G. Sotiropoulos, A.R. Seitz & P. Seriès, "Changing expectations about speed alters perceived motion direction". *Curr Biol*, 21(21), 2011, R883–R884.

²⁶ P.L. Smith & R. Ratcliff, "Psychology and neurobiology of simple decisions". *Trends Neurosci.* 27(3), 2004, 161-8.

²⁷ S. Denève, "Bayesian approach to decision making", *Handbook of reward and Decision making*, 2009.

²⁸ J. D. Roitman & M. N. Shadlen, "Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task". *J Neurosci.* 22(21), 2002, 9475-89.

< Figure 3 environ ici >

L'approche Bayésienne peut bien-sûr être appliquée aussi à des tâches plus complexes. Les travaux du laboratoire de Josh Tenenbaum et collaborateurs²⁹, par exemple, suggèrent que l'approche Bayésienne peut permettre de répondre à des questions fondamentales sur l'acquisition des connaissances abstraites et aider à comprendre comment les connaissances peuvent guider l'apprentissage à partir de données incomplètes. Ces auteurs s'intéressent aux capacités cognitives qui semblent difficiles à atteindre pour les machines mais où les jeunes enfants excellent, par exemple l'apprentissage d'une tâche ou d'un concept à partir de seulement quelques exemples et leur généralisation, ou la compréhension qui nous semble de "bon sens" du monde physique ou des relations sociales. Ils montrent que l'acquisition de ces capacités cognitives peut être décrite comme un processus d'extraction des structures présentes dans les informations reçues et peut être modélisée à l'aide de modèles probabilistes hiérarchiques utilisant des représentations de connaissance classiques dans le domaine de l'intelligence artificielle, tels que des graphes, grammaires ou schémas, calculs de prédicats et programmes fonctionnels. Ils proposent que ces modèles puissent être intégrés pour former une architecture cognitive unifiée, en utilisant des langages de programmation probabiliste.

Psychiatrie Computationnelle.

Les modèles Bayésiens ont enfin récemment fait leur apparition dans le domaine de la psychiatrie où ils semblent particulièrement prometteurs³⁰. Dans ce domaine, les modèles Bayésiens permettent de quantifier les différences de performances entre différents groupes (par exemple, sains vs malades) et potentiellement d'identifier si ces différences proviennent de l'utilisation de modèles internes ou de stratégies d'apprentissage ou de décision différents.

Dans l'étude de la schizophrénie par exemple, des travaux récents suggèrent que les patients pourraient présenter des déficits en terme d'inférence probabiliste. Une tâche couramment étudiée pour étudier cet effet est la tâche dite "des urnes" où les participants doivent deviner de quelle urne provient une perle tirée au hasard. Dans la version originale de la tâche, une urne contient 85% de perles rouges et 15% de perles noires, alors que l'autre urne contient 15% de perles rouges et 85% de perles noires. L'on tire au hasard une perle à la fois et l'on demande aux participants d'indiquer quand ils ont reçu suffisamment d'informations pour décider de quelle urne les perles ont été tirées. Les patients schizophrènes sont plus susceptibles que les contrôles d'arrêter leur décision après un petit nombre d'observations (1 ou 2 tirages) et d'avoir un sentiment de certitude après seulement un tirage - une tendance à tirer des conclusions hâtives³¹ qui pourrait être centrale pour expliquer les délusions et délires paranoïaques. En utilisant les modèles de décision présentés ci-dessus pour les tâches de choix forcé alternatif (Figure 3B) et comparant les performances des participants au modèle Bayésien optimal à des variantes sous-optimales, il est possible de mieux caractériser le déficit chez les patients et son origine potentielle. Il semble ainsi que les performances des patients seraient davantage liées à la présence d'un seuil de décision plus bas³² ou plus variable que chez les sujets contrôles³³, plutôt qu'une surestimation du coût associé à la collecte d'information supplémentaires.

Une idée courante dans le domaine de la psychiatrie est que les modèles internes de l'environnement utilisés par les patients, et en particulier les distributions *a priori*, pourraient être différents de ceux des sujets

²⁹ J.B. Tenenbaum, C. Kemp, T.L. Griffiths & N.D. Goodman, "How to grow a mind: statistics, structure and abstraction". *Science*. 331(6022), 2011, 1279-85.

³⁰ P.R. Montague, R.J. Dolan, K.J. Friston & P. Dayan, "Computational Psychiatry". *Trends Cogn Sci*, 16(1), 2012, 72-80.

³¹ appelé "jumping to conclusions" en anglais.

³² B. B. Averbeck, S. Evans, V. Chouhan, E. Bristow & S.S. Shergill, "Probabilistic learning and inference in schizophrenia". *Schizophr Res*, 127(1-3), 2011, 115-22.

³³ M. Moutoussis, R.P. Bentall, W. El-Deredy & P. Dayan, "Bayesian modelling of "Jumping-to-conclusion" bias in delusional patients". *Cogn Neuropsychiatry*, 16(5), 2011, 422-47.

sains. Dans l'étude de la schizophrénie, par exemple, il a été proposé que les symptômes dits "positifs" (hallucinations et délusions) puissent être décrits en termes d'un déséquilibre entre les informations ascendantes (vraisemblance) et descendantes (*a priori*) dans un système Bayésien hiérarchique perturbé³⁴. Dans l'étude de l'autisme, il a récemment été proposé que les distributions *a priori* pourraient jouer un rôle trop faible par rapport aux entrées sensorielles (vraisemblance)³⁵. Cela pourrait expliquer que les patients puissent se sentir submergés par un monde perçu comme étant "trop réel".

Dans l'étude de la dépression de la même manière, les modèles probabilistes ont fait leur apparition. Huys et Dayan (2009)³⁶ se sont intéressés par exemple au phénomène d'impuissance apprise ("learned helplessness"), l'état psychologique où un animal ou une personne a appris que son comportement n'avait aucune influence sur le cours des choses et se résigne et devient apathique, un état de désespoir qui est souvent considéré comme un modèle de la dépression. Ces auteurs montrent que ce phénomène peut être caractérisé en terme de distribution *a priori* sur la contrôlabilité. Nous avons montré de même que l'optimisme (qui fait défaut dans la dépression) pouvait dans certains cas être modélisé en terme de distribution *a priori* sur la probabilité des récompenses futures³⁷ et peut être quantifié précisément chez les individus grâce à la modélisation Bayésienne.

Le succès de modèles Bayésiens.

En conclusion, dans le domaine de la perception comme dans celui de la prise de décision, l'approche Bayésienne paraît particulièrement adaptée pour décrire les performances humaines. Elle offre une description élégante et parcimonieuse du comportement. Les modèles Bayésiens ont de plus l'avantage d'être transparents quant aux hypothèses sur lesquels ils sont bâtis et mettent l'accent sur l'aspect fonctionnel des processus cognitifs. Comme on l'a vu, les données recueillies sont en général compatibles avec l'hypothèse Bayésienne dans le sens où elles montrent que le cerveau prend en compte l'incertitude associée aux signaux qu'il reçoit pour les intégrer de manière (quasi-)optimale. Le cerveau semble aussi prendre en compte certaines connaissances *a priori*, qui sont constamment mis à jour, sont qualitativement compatibles avec les statistiques du monde et peuvent être décrites chez le sujet individuel en terme de probabilité de distribution *a priori*. De plus, les modèles Bayésiens fournissent des outils précieux pour quantifier des différences de comportements entre individus et pourraient éventuellement aider au diagnostic lorsque les capacités sont détériorées, comme par exemple dans le domaine de la psychiatrie.

3. Substrat neuronal du calcul Bayésien.

Si les modèles Bayésiens semblent particulièrement utiles pour décrire les performances perceptives ou liées à la prise de décision, leur relation avec le substrat neuronal est un sujet controversé³⁸. Pour

³⁴ P.C. Fletcher & C.D. Frith, "Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia". *Nat Rev Neurosci.* 10(1), 2009:48-58.

³⁵ E. Pellicano & D. Burr, "When the world becomes 'too real' a Bayesian explanation of autistic perception". *Trends Cogn Sci.* 16(10), 2012, 504-10; mais voir aussi K.J. Friston, R. Lawson & C.D. Frith, "On hyperpriors and hypopriors: comment on Pellicano and Burr". *Trends Cogn Sci.* 17(1), 2013:1.

³⁶ Q.J. Huys & P. Dayan, "A Bayesian formulation of behavioural control". *Cognition.* 113(3), 2009, 314-28.

³⁷ A. Stankevicius, Q. Huys, A. Kalra & P. Seriès, « Optimism as a prior on the likelihood of future reward », en révision.

³⁸ voir par exemple : M. Colombo & P. Seriès, "Bayes in the brain – on Bayesian modelling in neuroscience". *British Journal for the Philosophy of Science*, 63, 2012, 697–723; J.S. Bowers & C.J. Davis, "Bayesian just-so stories in psychology and neuroscience". *Psychol. Bull.* 138(3), 2012, 389-414; J.X. O'Reilly, S. Jbabdi & T.E. Behrens, "How can a Bayesian approach inform neuroscience?". *Eur J Neurosci*, 35(7), 2102, 1169-79.

beaucoup d'auteurs, il existe un fossé difficilement franchissable entre les modèles Bayésiens qui rendent compte des processus cognitifs à un niveau computationnel et la compréhension des mécanismes neurobiologiques qui pourraient sous-tendre ces calculs. Parce qu'ils décrivent un niveau d'analyse différent, il n'est ainsi pas clair que les modèles Bayésiens puissent être prédictifs au niveau biologique (voir aussi §4). Néanmoins, il existe de nombreuses propositions concernant la manière dont l'inférence Bayésienne pourrait être implémentée. Ces propositions cherchent en général à montrer que "l'hypothèse du codage Bayésien", c.-à-d. l'idée selon laquelle le cerveau pourrait représenter et manipuler des densités de probabilités, n'est pas incompatible avec ce que l'on sait sur le fonctionnement du cerveau³⁹.

Certains modèles proposent par exemple que l'activité r (fréquence de décharge) des neurones sélectifs à une caractéristique particulière s (par exemple l'orientation d'un objet visuel) pourrait être directement proportionnelle à la probabilité $p(s|r)$ ou au logarithme de la probabilité que cette caractéristique soit présente dans le champ récepteur du neurone ou que cette caractéristique s prenne une valeur donnée. Si la variable encodée est binaire, une autre proposition est que l'activité neuronale reflète le rapport de vraisemblance logarithmique ($r \propto \log[p(s=1)/p(s=0)]$). D'autres modèles proposent que les probabilités soient exprimées de manière implicite comme la somme pondérée de fonctions préexistantes (appelées « fonctions de base », par exemple les courbes d'accord des neurones). L'activité neuronale représenterait les coefficients de cette somme pondérée. Plus précisément, le logarithme de la probabilité $p(s|r)$ serait représenté implicitement comme:

$$\log p(s|\mathbf{r}) = \sum_i r_i h_i(s) + \text{constante}$$

Où les r_i représentent les activités des neurones i , et $h_i(s)$ sont les fonctions de base associées aux neurones i , plus une constante, nécessaire pour la normalisation. Quand les fonctions de base sont dérivées de la fonction de vraisemblance $\log p(r|s)$ (qui est en général décrit comme une fonction des courbes d'accord et du bruit), le modèle résultant est appelé un "modèle de codage par population probabiliste linéaire". Ma et al (2006)⁴⁰ ont montré que cette forme de codage était naturellement obtenue si la variabilité des réponses neuronales est proche d'un processus de Poisson -- ce qui semble être le cas dans le cortex. Ces auteurs ont de plus montré qu'un tel code était particulièrement adapté pour le calcul d'inférences optimales dans les situations d'intégration multi-sensorielle. En effet, celui-ci nécessite une opération multiplicative au niveau des distributions *a posteriori* (Eq. 2), mais il peut-être réalisée en effectuant une simple addition des réponses des populations neuronales avec ce modèle.

De manière alternative, les probabilités pourraient être représentées à tout moment par des échantillons tirés de cette probabilité⁴¹. Selon ce modèle, chaque neurone représenterait une caractéristique, ou "cause" des entrées sensorielles que le cerveau cherche à extraire. Les probabilités seraient représentées par la variabilité des réponses: les neurones effectueraient un tirage aléatoire de leur nombre de potentiels d'action à partir de la probabilité de distribution correspondant à la variable représentée. Ainsi par exemple, un neurone dont le taux de décharge instantané oscille autour d'une moyenne de 50 potentiels d'action avec peu de variabilité au cours du temps indiquerait à une forte probabilité que cette variable soit présente. Réciproquement, un neurone présentant une grande variabilité au cours du temps représenterait une forte incertitude sur l'absence ou la présence de la variable encodée. Cette représentation nécessite que de multiples échantillons (soit une séquence de mesures du taux de décharge instantané) soient recueillis au cours du temps pour permettre une estimation satisfaisante de la distribution représentée, mais peut en principe représenter n'importe quelle distribution. Ce modèle semble particulièrement bien adapté pour représenter des distributions multidimensionnelles, pour l'acquisition de ces distributions et pour certaines opérations comme l'opération de marginalisation.

³⁹ Pour une revue récente voir: A. Pouget, J.M. Beck, W.J. Ma & P.E. Latham, "Probabilistic brains: knowns and unknowns." Nat. Neuroscience, 16(9), 2013, 1170-8.

⁴⁰ W.J. Ma, J.M. Beck, P.E. Latham & A. Pouget, "Bayesian inference with probabilistic population codes". Nat Neurosci, 9(11), 2006, 1432-8.

⁴¹ J. Fiser, P. Berkes, G. Orbán & M. Lengyel, "Statistically optimal perception and learning: from behaviour to neural representations". Trends Cogn Sci. 14(3), 2010, 119-30.

D'autres laboratoires ont cherché à comprendre plus particulièrement comment les distributions de probabilité *a priori* pouvaient être implémentées dans le cerveau. Il existe des données suggérant que les fonctions de vraisemblance et les distributions *a priori* seraient encodées séparément dans le cerveau⁴² mais l'on sait peu de choses sur l'implémentation de ces probabilités. Dans le domaine sensoriel par exemple, il n'est pas clair si les distributions *a priori* correspondent à des signaux descendants provenant d'aires corticales supérieures, si elles résident entièrement dans ces aires corticales supérieures, ou bien au contraire si elles peuvent être formées et maintenues dans le cortex sensoriel lui-même. Plusieurs propositions existent pour ce qui est de l'implémentation des distributions de probabilité *a priori* (qui ne sont pas forcément toutes mutuellement exclusives). Il a ainsi été proposé que les distributions *a priori* correspondent à une augmentation de l'activité des neurones sélectifs aux caractéristiques du stimulus attendu, d'une manière similaire aux phénomènes attentionnels. Dans d'autres cas, les distributions *a priori* pourraient correspondre à une réduction des activités, comme dans les processus d'adaptation⁴³. Une autre proposition est que les probabilités *a priori* découlent implicitement de la sélectivité des neurones et de l'inhomogénéité de leurs caractéristiques préférées⁴⁴. Ainsi par exemple, l'*a priori* pour les orientations cardinales pourrait résulter du fait qu'il existe davantage de neurones préférant les orientations cardinales dans le cortex visuel primaire. Une autre proposition intéressante est que les distributions *a priori* pourraient correspondre à un changement de l'activité de base⁴⁵ où les neurones sélectifs aux caractéristiques attendues augmenteraient leur activité spontanée. Le cerveau serait ainsi constamment dans un état d'alerte où l'activité spontanée correspondrait à une prédiction du stimulus futur le plus probable.

4. Le cerveau est-il Bayésien? Critiques et Questions Ouvertes

L'approche Bayésienne en neuroscience est à la fois extrêmement en vogue et la cible de violentes critiques. Bowers et Davis (2012)⁴⁶ écrivent par exemple:

“notre argument principal est que la modélisation Bayésienne, dans la pratique comme dans le principe, est une approche fourvoyée⁴⁷ pour l'étude de l'esprit et du cerveau.”

Les critiques concernent plusieurs points que nous détaillons ici. La critique la plus véhémente est liée au fait que l'approche Bayésienne semble difficilement falsifiable. L'approche Bayésienne semble si flexible qu'il est toujours possible d'ajuster un modèle Bayésien aux données recueillies, si l'on choisit la probabilité *a priori* et la fonction de coût du modèle pour rendre compte des données. Un autre élément dans ce débat est lié à une certaine confusion sur la nature de l'argument concernant l'optimalité. En effet, de nombreuses publications dans ce domaine ont utilisé le terme séducteur d'optimalité dans leurs titres ou résumé – et ce n'est que récemment que la distinction a été faite entre les termes de « Bayésien » et « optimal ». Cependant, l'approche Bayésienne ne suppose pas que le système nerveux est construit de manière optimale, ni que le modèle interne utilisé dans le raisonnement est une copie exacte du modèle externe du monde (ou bien il faut que cela soit démontré), mais que la *combinaison* des signaux impliqués dans le calcul (qui résultent potentiellement de processus sous-optimaux) est optimale dans le sens où elle est

⁴² U.R. Beierholm, S.R. Quartz & L. Shams, “Bayesian priors are encoded independently from likelihoods in human multi- sensory perception”. *J Vis*, 9(5), 2009, 23.1–23.9 ; I. Vilares, J.D. Howard, H.L. Fernandes, J.A. Gottfried & K.P. Kording. “Differential representations of prior and likelihood uncertainty in the human brain. *Curr Biol*”, 22(18), 2012, 1641–1648.

⁴³ C. Summerfield & T. Egner, « Expectation (and attention) in visual cognition ». *Trends Cogn Sci*, 13(9), 2012, 403–409.

⁴⁴ A.R. Girshick, M.S. Landy & E.P. Simoncelli, « Cardinal rules: visual orientation perception reflects knowledge of environmental statistics ». *Nat Neurosci*, 14(7), 2011, 926–932.
B.J. Fischer B.J. & J.L. Pena, « Owl's behavior and neural representation predicted by Bayesian inference ». *Nat Neuro* 3; 14(8), 2011, 1061-6.

⁴⁵ P. Berkes, G. Orban, M. Lengyel & J. Fiser. « Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment ». *Science*, 331(6013), 2011, 83–87.

⁴⁶ J.S. Bowers & C.J. Davis, « Bayesian just-so stories in psychology and neuroscience », *Psychol. Bull.* 138(3), 2012, 389-414; voir aussi M. Jones & B.C. Love, « Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition ». *Behav. Brain Sci.* 34(4), 2011, 169-88.

⁴⁷ en anglais: “misguided”.

réalisée selon la formule de Bayes. De plus, devant la complexité et la multi-dimensionnalité des situations auxquelles le cerveau doit faire face, tous les auteurs semblent s'accorder à dire que seule une approximation du calcul Bayésien est possible. Dans certains cas, même le calcul d'une approximation paraît d'une complexité insurmontable⁴⁸. Sur le plan fonctionnel, l'impact de ces approximations n'est pas clair et quelle approximation peut-être encore qualifiée de Bayésienne? Comment distinguer clairement ce qui est Bayésien de ce qui ne l'est pas, dans ce contexte? Selon Bowers et Davis (2012), un autre élément contestable est que les modèles Bayésiens sont rarement comparés à d'autres modèles non-Bayésiens, par exemple aux modèles connexionnistes -- il ne semble donc pas y avoir de test concernant une hypothèse alternative. Pour d'autres auteurs⁴⁹, le problème est qu'en général il existe non pas un modèle mais une famille de modèles Bayésiens pour décrire une tâche donnée (selon les choix et paramètres utilisés pour la vraisemblance, distributions *a priori*, fonction de coût etc.). Souvent, un seul de ces modèles est choisi pour ajuster les données et le problème de la sélection de ce modèle parmi les autres n'est pas clairement discuté. De la même manière, Marcus et Davis (2012) suggèrent que la littérature est biaisée par le fait que les tâches étudiées et publiées sont celles où les performances humaines peuvent être décrites comme étant quasi-optimales, alors que les résultats ou situations qui ne sont pas facilement décrits par des modèles Bayésiens ne sont pas discutés. Enfin, comme on l'a vu, un sujet de controverses est le fait que l'approche Bayésienne semble faire peu de prédictions sur le plan biologique, et donc a peu de support expérimental à ce niveau.

Ce débat est toujours ouvert, mais certaines réponses ont été proposées. Contre la critique relative au manque de falsifiabilité, un argument proposé par Griffith et coll. (2012)⁵⁰ est que l'approche Bayésienne doit être considérée comme un cadre théorique, au même titre que l'approche connexionniste ou les architectures cognitives symboliques, comme ACT-R. Pour ces auteurs, un cadre théorique ne peut être invalidé, par définition; seuls les modèles qu'il produit peuvent l'être. L'évaluation d'un cadre théorique est liée à sa capacité à produire des modèles satisfaisants. Griffith et coll. argumentent ainsi que si l'approche Bayésienne ne peut être invalidée, les instanciations de cette approche par des modèles Bayésiens spécifiques peuvent l'être. Un autre argument est lié au fait que les modèles Bayésiens sont avant tout des modèles fonctionnels. David Marr (1982)⁵¹ avait proposé que les calculs neuronaux puissent se comprendre à différents niveaux d'analyse, largement indépendants. Il distingue le niveau computationnel (quel est le but du calcul? pourquoi est-il justifié?), le niveau algorithmique (quelle est la représentation des entrées et sorties, et quel algorithme peut décrire cette transformation?) et le niveau implémentatif (comment ces représentations et algorithmes sont-ils implémentés physiquement?). Les modèles Bayésiens sont des modèles qui correspondent au niveau computationnel selon la classification de Marr et non au niveau algorithmique ou implémentatif. Comme l'écrivent Jacobs & Kruschke (2010)⁵²: "Les modèles Bayésiens n'ont pas pour but de fournir une description mécanistique de la cognition". Il n'est donc pas justifié de comparer les modèles Bayésiens à des modèles décrits à un niveau d'analyse différent. L'hypothèse du cerveau Bayésien (contrairement à l'hypothèse du codage Bayésien) ne serait donc pas incompatible avec des modèles mécanistiques qui seraient fondés sur de simples heuristiques. En effet, Griffith et coll. écrivent:

"Le fait d'avoir quelque chose d'intrinsèquement Bayésien n'est aucunement une nécessité pour les algorithmes approximaient l'inférence Bayésienne."⁵³

Dans ce contexte, il existe une certaine confusion concernant les implications de l'hypothèse du cerveau Bayésien pour le substrat neuronal et quels sont les auteurs qui défendent réellement l'hypothèse du codage Bayésien -- et sous quelle forme.

⁴⁸ J. Kwisthout, T. Wareham & I. van Rooij, "Bayesian intractability is not an ailment that approximation can cure". *Cogn Sci.* 35(5), 2011, 779-84.

⁴⁹ G.F. Marcus & E. Davis, "How Robust Are Probabilistic Models of Higher-Level Cognition?", *Psychol. Sci.* 2013, oct 1.

⁵⁰ T.L. Griffiths, N. Chater, D. Norris & A. Pouget, "How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012)." *Psychol Bull.* 2012 May;138(3):415-22.

⁵¹ D. Marr, "Vision: A Computational Approach". San Francisco, Freeman & Co, 1982.

⁵² R. A. Jacobs & J.K. Kruschke, "Bayesian learning theory applied to human cognition" *Wiley Interdisciplinary Reviews: Cognitive Science*, 2, 210, 8-21.

⁵³ "There needs to be nothing intrinsically Bayesian about algorithms that approximate Bayesian inference."

Conclusion

En conclusion, les modèles Bayésiens connaissent un succès indéniable pour l'étude des performances humaines et animales. Dans ce domaine, ils offrent des outils efficaces et précieux pour décrire le comportement, découvrir les stratégies sous-jacentes, et quantifier les différences entre individus. Leur rôle pour ce qui est de comprendre les mécanismes neurobiologiques associés est plus controversé. Ces dernières années ont été marquées par de nombreux débats concernant les termes utilisés dans ce domaine, en particulier les notions de falsifiabilité et d'optimalité, et les implications des résultats obtenus pour la compréhension du cerveau. Pour certains, l'approche Bayésienne ne représente qu'un nouveau vocabulaire, efficace, mais relativement creux, qui ne décrit que des faits déjà connus. Pour d'autre, l'approche Bayésienne permettrait de fournir un modèle unifié de la cognition humaine. Ces débats paraissent bénéfiques et nécessaires pour que ce domaine de recherche atteigne un état de maturité et clarifie l'importance de ses résultats. Une chose est sûre: l'approche Bayésienne est aujourd'hui une source d'inspiration extrêmement riche pour l'exploration de nouveaux paradigmes en neurosciences, analyses et questions expérimentales, que ce soit sur le plan théorique, comportemental ou physiologique.

Remerciements

Mes sincères remerciements vont à Vincent Valton et Freddy Grimault pour leur lecture attentive d'une version préliminaire de ce chapitre et pour leurs conseils.

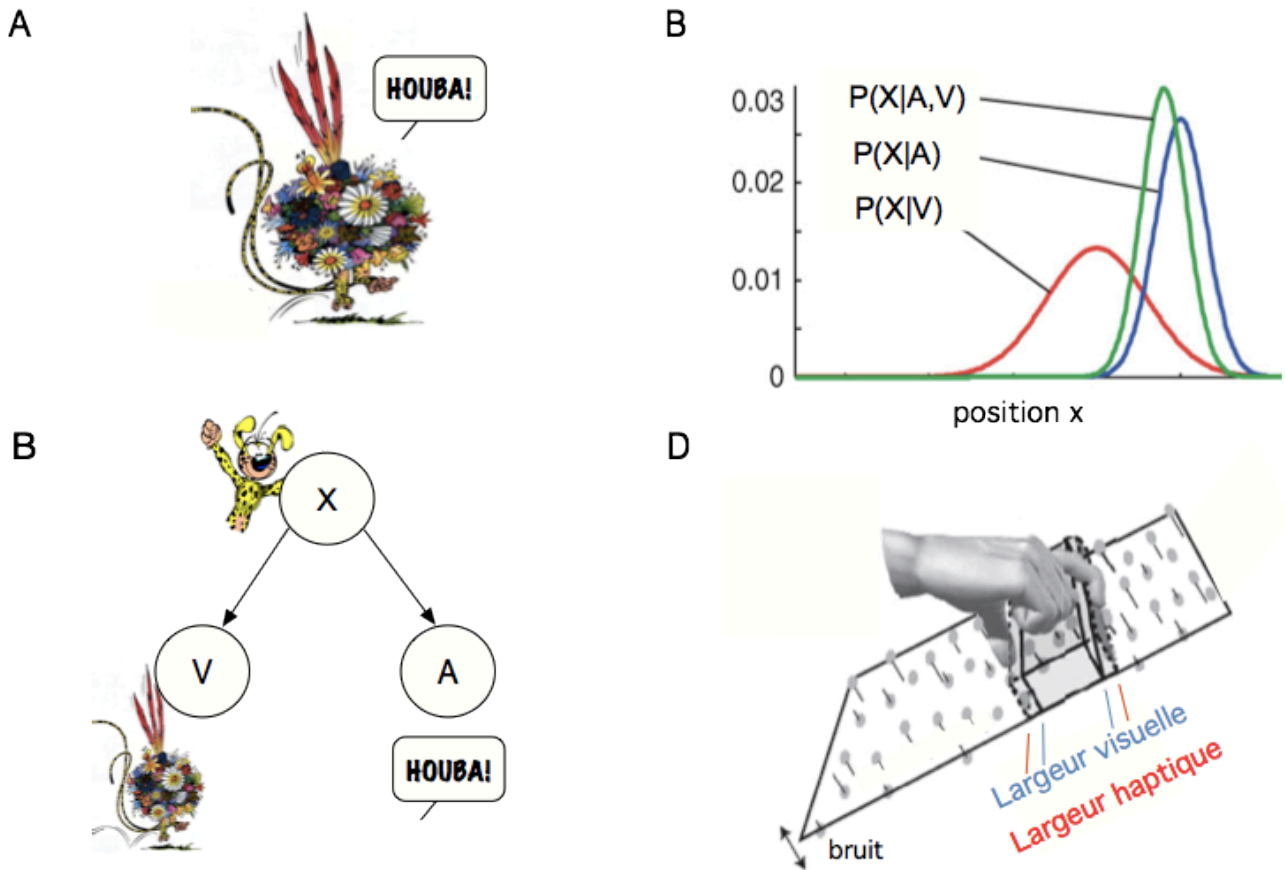
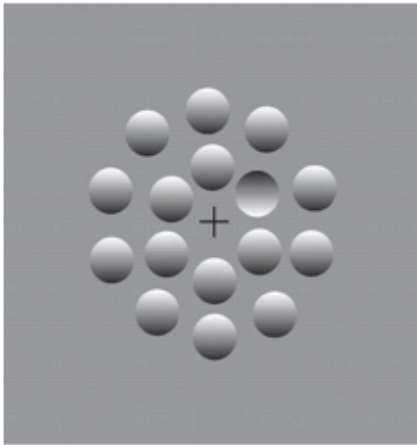


Figure 1: Intégration multimodale. **A.** Exemple de l'observation indirecte d'un prédateur potentiel. L'image visuelle offre une information ambiguë relative à la position de l'animal, mais l'on peut aussi entendre son cri: comment combiner ces deux informations pour déduire la position de l'animal? **B.** Modèle graphique correspondant à la situation de A. La variable X (position de l'animal) n'est pas observée mais elle produit deux variables observées: le mouvement d'un massif de fleurs qui fournit une indication visuelle, V, et le cri de l'animal, qui fournit une information auditive, A. **C.** Le modèle Bayésien indique comment ces informations doivent être combinées pour une estimation optimale. La distribution *a posteriori* $P(X|A,V)$ est influencée davantage par la fonction de vraisemblance la plus fiable (ici, l'information auditive $P(X|A)$). De plus, la précision de l'estimation finale (qui correspond à la largeur de la distribution *a posteriori*) est meilleure que celle qui serait obtenue pour chacune des deux modalités prise de manière isolée. **D.** Expérience de Ernst & Banks (2002) cherchant à tester le modèle Bayésien. Ici, les participants doivent estimer la largeur d'une arête en relief à l'aide d'informations visuelles (plus ou moins bruitées) et d'informations tactiles. (A et B sont inspirés de Vilares & Kording (2011) et des dessins de Franquin; C et D sont adaptés de Pouget et coll. (2013)).

A



B



C

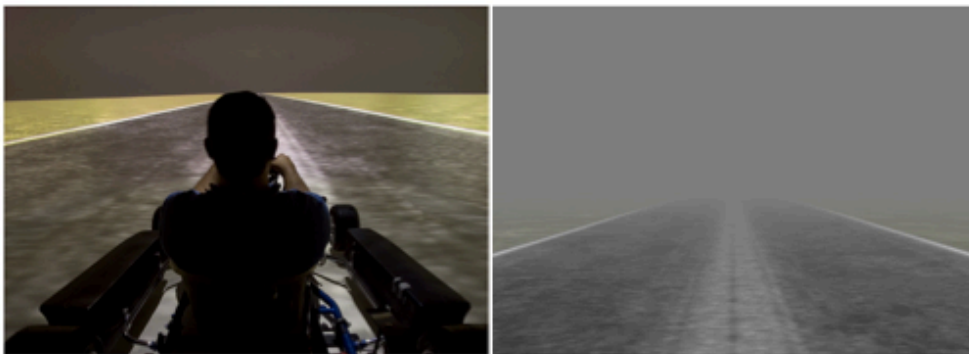


Figure 2. Illusions visuelles et distributions *a priori*. **A.** Le cerveau semble fonctionner avec l'hypothèse implicite que la lumière provient du haut. Ces formes sont-elles des bosses ou des creux? Percevoir un creux au milieu de bosses est compatible avec l'hypothèse selon laquelle la lumière provient d'en haut. Pour vous en convaincre, retournez la page et vous percevrez un percept opposé. **B.** Chambre de Ames. Les personnages sont perçus comme s'ils avaient des tailles très différentes. Cet effet est lié au fait que le cerveau interprète cette image comme si la géométrie de la pièce était cubique. En réalité, elle est trapézoïdale et le personnage qui semble tout petit est beaucoup plus éloigné qu'il ne paraît. **C.** De nombreuses études montrent que le cerveau tend à sous-estimer la vitesse des objets faiblement contrastés. Cet effet reflèterait l'utilisation de distributions *a priori* favorisant les vitesses lentes et pourrait expliquer pourquoi certains conducteurs ont tendance à conduire trop vite dans le brouillard. Figure C adaptée de Pretto et al (2012)⁵⁴, qui explorent ces effets en utilisant un simulateur de conduite.

⁵⁴ P. Pretto, J.P. Bresciani, G. Rainer & H.H. Bulthoff, « Foggy perception slows us down », *Elife*, 2012, 1:e00031.

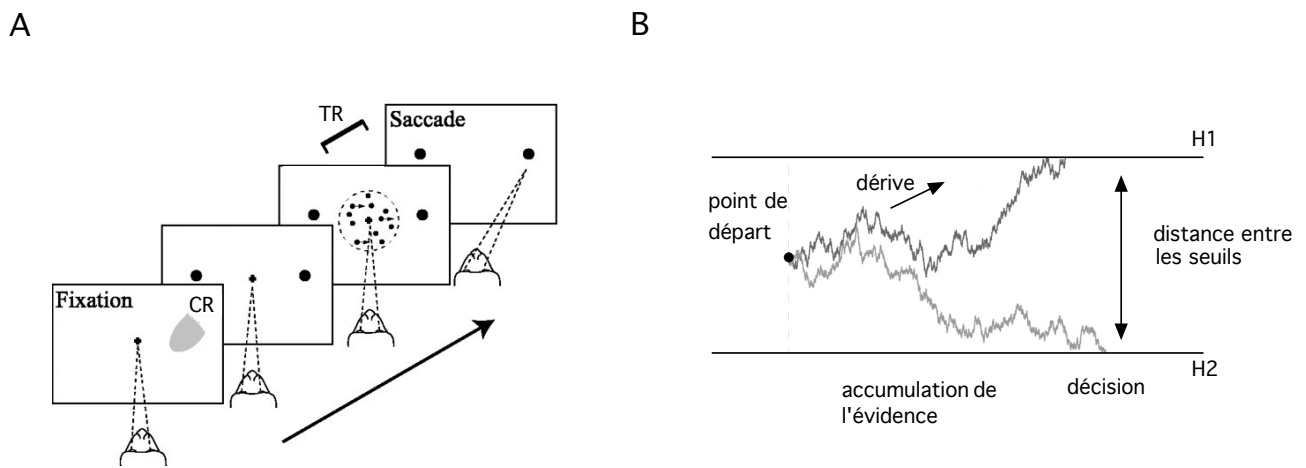


Figure 3. Modèle de prise de décision. **A.** Exemple d'une tâche de choix forcé alternatif. Ici, l'animal doit décider si les points aléatoires se déplacent vers la gauche (H1) ou vers la droite (H2) et indiquer sa réponse en effectuant une saccade visuelle vers la gauche ou la droite. Les taux de réussite et temps de réaction (TR) sont mesurés. **B.** Les performances comportementales humaines peuvent être modélisées de manière satisfaisante par un processus de diffusion d'une particule dont la dérive vers les seuils de décision correspondant aux deux hypothèses (H1 et H2) dépend des informations reçues (l'évidence). Dans ce modèle, les taux d'erreurs et le temps de décision dépendent de la position initiale de la particule et de la distance entre les seuils. Ce modèle peut être exprimé comme un modèle Bayésien, similaire au test séquentiel du rapport de vraisemblance. A est adapté de Roitman & Shadlen (2002).