Cognition *in eco*, cognition *in vitro*

Mesure, explication et
naturalisation des
comportements humains
(Travail en cours)

Avel Guénin—Carlut
KAIROS Research

avel@kairos-research.org

Abstract

L'explication des comportements humains par la psychologie cognitive suppose l'existence d'un lien direct entre les régularités comportementales mesurées *in vitro* et celles qui s'expriment *in eco*. Au contraire, les neurosciences cognitives s'appuie sur une modélisation causale explicite des dynamiques qui sous-tendent la cognition, et considèrent les comportements humains comme une expression contextuelle de leur architecture cognitive. Nous exposerons ici la divergence méthodologique fondamentale entre ces deux démarches, qui encouragent respectivement une réduction des comportements humains à un ensemble d'effets psychologiques individualisés et une intégration de la cognition dans son contexte biologique et culturel. Nous montrerons que la capacité des systèmes cognitifs à produire activement les contraintes qui sous-tendent leur propre activité invalide la première approche, et proposerons le cadre formel et conceptuel de l'*Active Inference* comme un candidat pour intégrer l'étude des différentes échelles de comportement qui constituent la cognition. Nous espérons ainsi mettre en lumière la nécessité absolue d'une démarche naturaliste et intégrative dans l'étude de la cognition, et plus généralement dans l'explication des comportements humains.

The explanation of human behavior by cognitive psychology assumes a direct link between behavioral regularities measured *in vitro* and those expressed *in eco*. In contrast, cognitive neuroscience relies on an explicit causal modeling of the dynamics underlying cognition, and considers human behaviors as a contextual expression of their cognitive architecture. We will expose here the fundamental methodological divergence between these two approaches, which respectively encourage a reduction of human behavior to a set of individualized psychological effects and an integration of cognition in its biological and cultural context. We will show that the ability of cognitive systems to actively produce the constraints that underlie their own activity invalidates the first approach, and propose the formal and conceptual framework of *Active Inference* as a candidate for integrating the study of the different scales of behavior that constitute cognition. We thus hope to highlight the absolute necessity of a naturalistic and integrative approach in the study of cognition, and more generally in the explanation of human behavior.

0 – Introduction

Durant les années 1940 et 1950, la volonté d'intégrer les différents champs disciplinaires en médecine a poussé la fondation philanthropique Macy à organiser une série de conférences ayant pour humble ambition de restaurer l'unité des sciences et, au passage, de poser les bases d'une « science générale du fonctionnement de l'esprit humain » ('Macy Conferences', 2021). Les conférences Macy donnèrent naissance à un programme de recherche transdisciplinaire nommé cybernétique, qui n'a pas survécu aux dites conférences mais dont les réflexions sur l'organisation biologique et l'information ont structuré de larges domaines de la recherche contemporaine, des approches formelles des systèmes dynamique à la psychiatrie computationnelle en passant par la théorie de l'information, la biologie systémique ou la sociologie relationnelle. Elles posèrent ainsi les bases sur lesquelles des psychologues et linguistes purent dans les années 1960 proposer d'ouvrir la « boîte noire » de l'esprit humain en élargissant les sciences comportementales à l'étude des processus sous-tendant la pensée humaine, et non aux seuls lois du comportement observable (comme le voulait l'école béhavioriste, alors perçue comme dominante) ('Cognitive Revolution', 2021; Wright & Bechtel, 2007). Cette étude s'appuya sur une analogie à la technologie informatique naissante, qui concevait l'esprit humain comme un logiciel implémenté dans l'architecture computationnelle du cerveau (Piccinini, 2010). Ainsi naquirent les sciences cognitives, un programme de recherche transdisciplinaire toujours actif se définissant principalement par son engagement méthodologique à étudier l'esprit humain comme système de traitement de l'information instancié par les dynamiques internes au cerveau.

Les sciences cognitives contemporaines semblent toutefois s'être détachées de leur programme transdisciplinaire historique. En effet, la littérature est aujourd'hui dominée par la psychologie cognitive, laquelle tend à s'abstraire des autres disciplines (anthropologie, linguistique, neuroscience, intelligence artificielle, et philosophie) supposées constituer les sciences cognitives (Núñez et al., 2019). Ce découplage traduit l'émergence d'un programme de recherche *cognitivist* centré sur l'explication des comportements par appel

aux résultats de la psychologie cognitive. Ce programme s'appuie directement sur le paradigme expérimental que la psychologie cognitive a hérité du béhaviorisme, lequel vise à inférer la manière dont l'architecture cognitive humaine représente et traite l'information en mesurant les régularités dans la réponse de sujets humains à des stimuli donnés en conditions contrôlées. Les processus cognitifs ainsi inférés en conditions expérimentales peuvent être projetés en conditions écologiques et ainsi expliquer les comportements humains. De ce fait, le programme cognitiviste d'explications des comportements suppose que les régularités comportementales produites par l'architecture cognitive humaines soient assez robustes pour motiver un lien fort entre la mesure des comportements *in vitro* et l'explication des comportements *in eco*. Cette idée est motivée par l'hypothèse du fonctionnalisme computationnel, descendante contemporaine de l'analogie entre cerveau et ordinateur, laquelle postule que l'architecture cognitive humaine constitue un système computationnel dont le fonctionnement est essentiellement autonome vis-à-vis de la structure causale sous-jacente. Il permet de fait de supposer que la cognition humaine est produit par des processus cognitifs dotés de lois stimulus-réponse régulière, indépendamment de l'architecture matérielle du cerveau, ainsi que des dynamiques le reliant au corps et à l'environnement dans lequel l'activité cognitive est intégrée.

Le programme de recherche cognitiviste est l'acteur central d'un programme de "naturalisation" des sciences sociales, visant essentiellement à refonder leurs résultats et leurs concepts de ces dernières par appel de ceux de la psychologie cognitive. Cette dernière aurait en effet montré l'invalidité de l'ensemble de présupposés anthropologiques développés au cours du temps par les sciences sociales (Kaufmann & Cordonier, 2011), dont le contenu devrait donc être naturalisé en le reformulant sur la base des entités postulées par la psychologie (c'est à dire des processus cognitifs définis par une loi stimulus-réponse régulière) (Sperber, 2011). En d'autres termes, nous devrions travailler à expliquer les faits sociaux sur la base des faits psychologiques établis par la démarche expérimentale de la psychologie cognitive, lesquels sont (souvent explicitement) tenues pour être des propriétés universelle de l'architecture cognitive humains. Cette démarche importe donc les présupposés méthodologique du programme cognitiviste, qui lui sont nécessaires pour motiver l'existence de "faits psychologiques" individués et mesurables par la moyen de la psychologie cognitive. Ce faisant, elle s'écarte d'une démarche d'explication des phénomènes observables par appel aux structures causales existant de le monde naturel,

et donc d'une démarche naturaliste similaire à celle suivie par les sciences du vivant contemporaines. En effet, ancrer l'explication des phénomènes psychologiques dans les phénomènes neurologiques impliquerait de resituer les régularités comportementales mesurées par la psychologie cognitive dans la structure causale de l'architecture cognitive et de son environnement, et menacerait donc la possibilité d'expliquer les comportements *in eco* par les régularités observables *in vitro*.

L'autre grand programme de recherche hérité des sciences cognitives historique, celui des neurosciences cognitives, comprend en effet les comportements humains comme une expression contextuelle de l'architecture cognitive. Ce dernier s'inscrit plus généralement dans la lignée des travaux initiés par la conférence de Macy sur l'organisation des systèmes complexes et le formalisme mathématique des systèmes dynamiques, laquelle a motivé une transition massive des sciences du vivant vers une stratégie d'explication causale-mécanique, s'appuyant sur la construction de modèles explicites visant à représenter l'organisation multi-échelle des systèmes vivants et leur situation dans leur contexte de fonctionnement (Bechtel, 2008; Mitchell, 2003). Les neurosciences cognitives inscrivent donc l'explication des comportements humains dans le cadre d'une démarche plus générale de modélisation explicite et systématique des structures causales qui produisent la cognition humaine (Boone & Piccinini, 2016). De ce fait, elle ne suppose pas l'existence de processus cognitifs définis par de simples relations stimulus-réponse observables en laboratoire, et ne permet pas d'expliquer les comportements *in eco* par la simple transposition des régularités comportementales mesurées *in vitro*. Cette approche de la cognition humaine implique au contraire une compréhension systématique des multiples échelles d'organisation matérielle du cerveau, mais aussi du rôle que joue leur couplage dynamique avec le reste de l'organisme et avec l'environnement dans lequel ce dernier évolue. Elle permet donc une intégration des sciences cognitives et sociales ne s'appuyant pas sur une réduction des unes aux autres, mais plutôt sur une modélisation explicite des structures et des processus causaux impliqués dans l'organisation sociale humaine et de leur rôle dans l'explication des comportements.

De fait, une approche intégrative est nécessaire pour étudier les processus physiques qui sous-tendent la cognition humaine. En effet, les systèmes cognitifs sont par construction des systèmes capables d'intégrer de l'information concernant leur environnement dans leur structure interne de sorte à maintenir un comportement adapté étant donné leur organisation

biologique et leur environnement. Au contraire des ordinateurs auquel les compare le programme cognitiviste, ils ne disposent pas d'un architecte ayant défini des relations input-output régulières aux différents processus qui sous-tendent leur activité, et ayant imposé une structure matérielle apte à maintenir ces relations. Ils s'appuient à la place sur un régime d'équilibre dynamique auto-organisé capable de produire une organisation transciente s'adaptant en temps réel aux régularités perçues de l'environnement, de sorte à maintenir la structure globale de l'organisme et sa situation au sein de sa niche écologique. De ce fait, les symmétries physiques constituant un système cognitif sont continuellement brisés et recrées par l'activité interne de ce dernier, dans un processus redéfinissant les régularités observables qui le caractérisent. Les effets psychologiques mesurables *in vitro* ne peuvent donc suffir à individuer les processus qui sous-tendent la cognition *in eco*, et les sciences cognitives doivent explicitement intégrer à leur explication des comportements humains l'étude du couplage entre l'agent cognitifs et son environnement ainsi que des processus socio-culturels qui produisent le sens qu'il projette sur son monde. Je propose ici de décrire le problème d'inférence qui se pose donc aux sciences cognitives, et de montrer comment l'apparente difficulté qu'il pose à une explication naturaliste des comportements humains est résolue par un paradigme intégratif décrivant la manière dont les systèmes cognitifs produisent leur réalité sensible : l'*Active Inference*.

Premièrement, nous exposerons d'abord pourquoi le programme de recherche cognitiviste suppose de fait un lien fort entre la mesure des comportements *in vitro* et l'explication des comportements *in eco*, et s'éloigne ainsi des engagements méthodologiques définissant le naturalisme. Deuxièmement, nous montrerons que le programme cognitiviste s'appuie implicitement sur l'hypothèse que les mécanismes cognitifs qui sous-tendent la cognition humaine sont individués par les comportements qu'ils produisent, et discuterons de la possibilité de justifier cette hypothèse par appel à la fonction ancestrale des mécanismes en question. Troisièmement, nous discuterons de la possibilités d'une "naturalisation" cognitiviste des comportements sociaux en exposant le débat contemporain entre deux théories concurrentes de l'évolution socioculturelle humaine défendant respectivement une réduction des dynamiques socio-culturelles aux processus cognitifs sous-jacents et une étude systématiquement de l'interaction entre les deux. Quatrièmement, nous discuterons de la possibilité de transposer les régularités comportementales mesurées *in vitro* pour expliquer les comportements *in eco* étant donné la capacité des systèmes cognitifs à

activement produire les contraintes structurelles qui déterminent leur propre activité, particulièrement via l'immersion participative dans des pratiques collective de *sensemaking* construisant des systèmes d'attente contextuels qui modulent la cognition humaine. Cinquièmement, nous proposerons l'intégration des différentes échelles d'analyse de la cognition dans le cadre d'une «science générale de l'esprit humain» dans le cadre formel et conceptuel de l'*Active Inference*. Cet article vise ainsi à montrer la nécessité d'une prise en compte explicite du rôle des dynamiques socio-culturelles dans l'explication des comportements humains, et l'opportunité de la réaliser dans le cadre d'une démarche explicitement naturaliste et intégrative dans l'étude de la cognition.

1 – Mesurer les dynamiques cognitives : la psychologie entre positivisme et naturalisme

L'approche cognitive s'est construite en opposition avec l'approche behavioriste, laquelle visait à établir les lois du comportement humain sur la base exclusive d'entités observables empiriquement, et rejetait donc catégoriquement tout appel aux états mentaux ou processus cognitifs. Elle a toutefois essentiellement repris son cadre expérimental : un *sujet* est soumis dans des conditions contrôlées à des *stimuli* auxquels on mesure ses *réponses*. Un traitement statistique permet ensuite de tirer des données brutes des régularités mesurables dans l'association stimulus-réponse des sujets, ou « effets » psychologiques. Ces effets (comme leur nom l'indique) sont généralement compris comme de simples produits de l'architecture cognitive humaine étant donné les conditions expérimentales, et sont donc l'objet (et non le sujet) des explications psychologiques (Cummins, 2000; Wright & Bechtel, 2007). Il demeure néanmoins courant de tenir de tels résultats comme des propriétés naturelles de la cognition humaine, lesquelles peuvent être généralisées en conditions écologiques, et ce malgré les importantes difficultés méthodologiques à établir la prévalence de tels « effets » - même au sein de la population étudiée (C. Spelman & McGann, 2013; C. P. Spelman & McGann, 2020). La psychologie cognitive partage donc avec la psychologie behavioriste un optimisme marqué concernant la validité écologique de ses résultats expérimentaux, au sens où la mesure de régularités comportementales *in vitro* est tenue pour s'appliquer aux comportements *in eco* et donc permettre de les expliquer. Nous allons montrer ici que cet optimisme traduit une récupération implicite des

stratégies explicatives behavioristes, et discuter des conséquences que cela a sur notre capacité à expliquer les comportements humains par les résultats de la psychologie cognitive.

L'économie comportementale fournit sans doute l'exemple le plus clair de la récupération d'une démarche behavioriste dans sa documentation des limites de la rationalité humaine (Kahneman, 2003) conçues comme un ensemble de «biais cognitifs», c'est à dire de déviations locales systématiques à des comportements normativement attendus étant donné les axiomes mathématiques définissant la rationalité. Par exemple, des sujets exposés en conditions expérimentales à des décisions en situation de risque (où ils sont exposés à un nombre limité d'options dont les différents résultats possibles sont connus et ont chacun une probabilité connue) n'adoptent pas un comportement maximisant l'espérance de leurs gains, mais semblent s'appuyer sur une perception distordue des probabilités extrêmes ainsi qu'une évaluation différente des gains et des pertes (Kahneman & Tversky, 1979). Ces biais sont ouvertement considérés comme s'appliquant aussi bien à des sujets en conditions expérimentales qu'en conditions écologiques, et certains chercheurs proposent par ailleurs d'exploiter en utilisant des suggestions (ou *nudges*) pour orienter les comportements humains à grande échelle (Thaler & Sunstein, 2009). Toutefois, l'interprétation de ces effets comme des limites à la rationalité s'appuie implicitement sur un accès optimal des sujets aux informations nécessaires à une décision optimale (Felin et al., 2017), alors même que les protocoles expérimentaux visant à les mesurer sont délibérément construits pour dissimuler les dites informations (Felin et al., 2019). Il apparaît donc que l'économie expérimentale n'accorde qu'un intérêt très limité à la documentation des processus cognitifs qui sous-tendent les effets qu'elle documente, au point où la validité de ses résultats puisse être assez largement remise en question.

Cet accent fort sur la caractérisation d'effets psychologiques en laboratoire, indépendamment de toute explication cognitive de leur occurrence, s'inscrit directement dans la foulée de la psychologie behavioriste. En effet, cette dernière souscrivait explicitement à un cadre méthodologique positiviste basé sur l'articulation de lois scientifiques universelles, elles-mêmes dérivées de l'expérience sensorielle directe des chercheurs (Graham, 2019; Wright & Bechtel, 2007). Cette démarche s'appuyait avant tout sur le rejet catégorique de toute affirmation non-vérifiable sur les états du monde (formellement, de tout énoncé synthétique *a priori*), conçus comme «métaphysiques» et donc dénués de sens (Graham, 2019, sec. 4.1). Elle impliquait également le

présupposé qu'il existe en effet des «lois psychologiques» régulières, lesquelles s'appliquent universellement aux dimensions observables du comportement humain, indépendamment de tout appel à des états mentaux inobservables dans la science psychologique. Les régularités comportementales mesurées en conditions expérimentales ne sont par conséquent pas comprises comme de simples produits contextuels de l'activité cognitive d'un ensemble d'individus donné, ni même des invariants de l'architecture cognitive humaine. Elle sont comprises comme l'expression directe de lois scientifiques proprement explicatives, caractéristiques de la nature humaine et donc *a priori* généralisables en conditions écologiques. Dans le cadre positiviste, établir la réalité de lois psychologiques ou expliquer un comportement se réduit donc à un problème de robustesse ou de réplicabilité d'un protocole expérimental capable de le susciter de manière mesurable. Il permettait donc à la psychologie expérimentale de revendiquer une forte autonomie disciplinaire, mais aussi un pouvoir explicatif quasiment illimité concernant les comportements humains.

Une telle démarche empiriste, s'appuyant avant tout sur l'hypothèse que la mesure *in vitro* des comportements suffit à en expliquer l'occurrence *in eco*, est largement partagée par la psychologie cognitive. En effet, l'approche cognitive s'est définie par l'analogie entre cerveau et ordinateur, formalisée *a posteriori* via la thèse du «fonctionnalisme computationnel», laquelle définit l'esprit comme organisation computationnelle du cerveau, à la manière d'un logiciel implémenté dans son architecture matérielle (Piccinini, 2010). Elle motive de ce fait un découpage entre «niveaux d'analyse» psychologiques (ou fonctionnelles) et neurologiques (ou computationnels) - laquelle est d'ailleurs explicitement revendiquée dans le découpage traditionnel entre niveaux d'analyse fonctionnel, algorithmique, et computationnel (Marr et al., 1979) - ou de manière plus extrême dans l'idée d'un «langage de la pensée» essentiellement détaché des dynamiques neurocognitives sous-jacentes (Fodor, 1975). L'approche cognitive suppose donc par construction une indépendance marquée de l'étude de la psychologie humaine, qu'elle conçoit (comme l'approche behavioriste avant elle) comme un niveau d'organisation autonome vis-à-vis de son architecture biologique ou de son contexte écologique. Bien que la psychologie cognitive contemporaine fasse appel aux concepts de «mécanisme» ou de «processus» cognitif, elle s'est détaché autour du temps des disciplines (neurosciences et intelligence artificielle) méthodologiquement capables de caractériser de telles entités (Núñez et al., 2019), et elle-même se contente de les approcher comme des effets psychologiques élaborés (comme l'illustre

l'exemple de l'économie comportementale). Malgré son opposition explicite à l'approche behavioriste et son accent sur la mesure des comportements observables, la psychologie cognitive partage donc par construction la même démarche d'explication des comportements humains.

L'approche de la psychologie cognitive est à contraster avec la stratégie d'explication causale-mécanique qui caractérise les neurosciences cognitives (Bechtel, 2008; Boone & Piccinini, 2016). Ces dernières, comme les autres sciences du vivant, vise typiquement à expliquer l'activité d'un système particulier en cartographiant ses propriétés structurelles. On identifie les parties et opérations élémentaires composant le système cible, on modélise comment leur organisation en produit le comportement global, puis enfin on replace ce modèle dans son contexte d'opération pour prédire et expliquer l'activité intégrée du système (Bechtel, 2009b). On peut illustrer cette démarche par l'exemple du renforcement, phénomène phare de la psychologie behavioriste, par lequel les sujets tendent à répéter des actions lorsqu'elles sont associées à des événements conçus comme plaisants. Les neurosciences cognitives ont identifié un ensemble de zones cérébrales impliquées dans ce phénomène (cortex posteromédial, cortex préfrontal, hippocampe, et jonctions temporopariétales), inféré le rôle que chacune pouvait avoir dans un processus de renforcement, puis en ont recomposé une structure vraisemblable reproduisant les équations dynamiques connues de l'apprentissage par renforcement (Dohmatob et al., 2020; Girard et al., 2008). Crucialement, les effets comportementaux y sont conçus comme le produit émergent de l'activité intégrée de l'architecture cognitive d'un sujet donné dans un contexte donné (Bechtel, 2009a; Bechtel & Abrahamsen, 2005). Rien ne présuppose qu'une telle architecture produise des régularités comportementales robustes mesurable *in vitro* dans le cadre d'un paradigme expérimental stimulus-réponse, et encore moins que la caractérisation de tels effets suffise à expliquer l'activité cognitive *in eco*.

La démarche explicative des neurosciences cognitives s'appuie sur la construction d'hypothèses explicites et plausibles concernant la structure causale sous-jacente (Boone & Piccinini, 2016; Craver, 2014). En d'autres termes, elle fait exclusivement appel à des entités physiques capables d'exercer une influence causale dans le monde naturel. Une telle démarche ne suppose pas nécessairement que toutes les entités supposées par les explications scientifiques doivent exister telle qu'elles sont décrites (Colombo et al., 2015), mais elle implique un engagement méthodologique à adopter une démarche naturaliste, compris comme une démarche explicative

invoquant un cadre ontologiquement cohérent avec la démarche propre aux sciences naturelles (Ramstead, 2015; Sperber, 2011). Le terme de naturalisme renvoie à la conception ontologique du terme (Ladyman et al., 2007; Papineau, 2020), c'est à dire le rejet de l'existence d'une causalité surnaturelle, ou plus la définition de la nature comme un domaine causalement clos où aucune entité ne peut exercer d'influence causale sans en faire elle-même partie (et donc être elle-même sujette à des influence causales externes) (Papineau, 2020). Puisque les «lois du comportement» et autres caractérisations purement fonctionnelles de la cognition ne font que supervenir sur la structure causale réelle qui sous-tend les dynamiques cognitives, elle ne peuvent donc être invoquées comme explications dans un cadre purement naturaliste (Mitchell, 2003, Chapter 5). De ce fait, nous pouvons désigner la démarche des neurosciences cognitives comme proprement *naturaliste*, mais celle de la psychologie cognitiviste doit plutôt être désignée (en référence à son assise dans le cadre méthodologique béhavioriste) comme *positiviste*.

Ces deux démarches partagent un mandat de «naturalisme méthodologique», c'est à dire d'unification des sciences entre elles et avec la philosophie (Papineau, 2020), mais divergent radicalement dans leur conception de l'unification. Les neurosciences cognitives se contraignent à considérer l'imbrication causale de différentes échelles de comportement (Boone & Piccinini, 2016), ou tout du moins leurs relations constitutives (Bechtel, 2008, 2009b, Chapter Four). Ces relations ne sont pas exhaustivement caractérisable à un niveau d'analyse ou une échelle de comportement en particulier, amenant donc nécessairement à une démarche transdisciplinaire d'intégration des connaissances scientifiques au sujet (Mitchell, 2003). Les approches expérimentales y sont donc recrutées pour cartographier la structure causale produisant le phénomène observé, et pas seulement pour établir une description purement fonctionnelle du comportement humain. Par exemple, les «neurosciences en deuxième personne» (Dumas et al., 2010, 2014) illustrent cette démarche par l'analyse exploratoire des dynamiques de synchronisation motrice et neuronale de sujets interagissant socialement en conditions expérimentales. Cette approche participe à une compréhension unifiée du monde naturel en s'intégrant dans une démarche de modélisation multi-échelle systématique du système cible, recrutant aussi bien des considérations ontogéniques, neurologiques ou sociales. Au contraire, l'appel implicite du cognitivisme à des lois comportementales lui permet de prétendre unifier les sciences en expliquant les «faits sociaux» par les propriétés de la nature humaine que constitueraient les «faits

psychologiques» révélés par la démarche expérimentale stimulus-réponse. Il s'inscrirait alors dans une démarche de « naturalisme » réductionniste au sens strict du terme, c'est à dire un programme visant à dériver les résultats d'une science des résultats d'une autre (van Riel & Van Gulick, 2019)).

Le rôle de la mesure *in vitro* dans l'explication des comportement *in eco* constitue donc une différence méthodologique fondamentale entre le programme de recherche cognitiviste et celui des neurosciences cognitives. Le premier tient de fait la mesure *in vitro* comme révélatrice d'effets psychologiques s'étendant par nature aux comportements *in eco*, adoptant donc une démarche positiviste qui lui permet de défendre une forte autonomie de la démarche expérimentale, mais également une explication cognitiviste des faits sociaux par les faits psychologiques mesurés en laboratoire. Au contraire, les neurosciences cognitives adoptent une démarche naturaliste qui ne conçoit la démarche expérimentale que comme un moyen d'inférer la structure causale qui sous-tend la cognition d'un sujet ou d'une population, et l'explication comme un exercice contextuel. Nous allons à présent nous intéresser à exposer les bases méthodologiques ainsi que les ambitions du programme cognitiviste en sciences sociales. Bien qu'il soit généralement désigné comme une forme de naturalisme (Andler, 2010; Sperber, 2011), nous allons voir que ses ambitions réductionnistes tout comme ses hypothèses méthodologiques sont incompatible avec une démarche naturaliste, et se rapproche plutôt d'une forme implicite de positivisme. Nous allons d'abord nous pencher sur la thèse du modularisme massif, laquelle sert à appuyer l'idée d'un lien direct entre mesure *in vitro* et explications des comportements *in eco* le rattache plus précisément au cadre méthodologique positiviste. Nous allons voir que cette dernière échoue non seulement à ancrer un tel lien dans une démarche naturaliste, mais révèle de plus une confusion problématique entre la structure physique de l'architecture cognitive et ses propriétés fonctionnelles mesurables.

2 – Explication cognitiviste des comportements & individuation fonctionnelle des mécanismes cognitifs

Nous avons vu ci-haut que l'explication des comportements écologiques par les régularités comportementales mesurées en conditions expérimentales dépend implicitement de l'hypothèse selon laquelle la démarche

expérimentale stimulus-réponse permet de révéler des propriétés naturelles de la cognition humaine. Nous avons également vu que, au sens d'une démarche d'explication naturaliste des dynamiques cognitives, ces propriétés se réduisent à la cartographie de la structure causale de l'architecture cognitive et des mécanismes qui la composent. De ce fait, les régularités exprimées en conditions expérimentales ne peuvent servir (dans le cadre d'une démarche naturaliste) que pour inférer les structures causales sous-jacentes, lesquelles peuvent éventuellement se trouver être explicatives de certains comportements observés en conditions écologiques. Vouloir expliquer les comportements *in eco* par les régularités observés *in vitro* dépend donc implicitement de l'hypothèse que les mécanismes cognitifs impliqués produisent une loi stimulus-réponse régulière, dont une caractérisation exhaustive suffirait à les décrire indépendamment de tout appel à leurs propriétés structurelles. Là où une telle condition d'*individuation fonctionnelle* est vérifiée, l'explication cognitive des comportements ne pose pas de soucis particuliers. En effet, on peut alors associer les régularités comportementales mesurées en conditions expérimentales à un mécanisme cognitif donné, qui produirait essentiellement la même régularité en conditions écologiques. De ce fait, pour pouvoir expliquer un comportement écologique par appel à des régularités mesurées en conditions expérimentales, l'explication cognitive des comportements se trouve confrontée au problème de devoir démontrer l'individuation fonctionnelle du mécanisme (ou de l'ensemble des mécanismes) produisant un comportement.

Hors, la thèse du «modularisme massif» (H. C. Barrett & Kurzban, 2006; Carruthers, 2006; Sperber, 2001) défend justement que la plupart des comportements sont produits par mécanismes cognitifs individuels fonctionnellement. L'argument central en est que l'architecture cognitive humaine est composée d'une agrégation de nombreux mécanismes cognitifs «modulaires», conçus comme des systèmes computationnels appliquant des opérations algorithmiques stéréotypées à une certaine catégorie d'inputs, lesquels ont été formés par la sélection naturelle en vertu de leur capacité à résoudre des problèmes écologiques particuliers dans un environnement ancestral dit «d'adaptation» ¹. Le modularisme massif prolonge et appuie de

1 «Un mécanisme psychologique évolué est un ensemble de processus à l'intérieur d'un organisme qui :

1. Existe sous la forme qu'il prend parce qu'il (ou d'autres mécanismes qui le produisent de manière fiable) a résolu un problème spécifique de survie ou de reproduction individuelle de manière récurrente au cours de l'histoire de

ce fait l'approche classiquement associée à la psychologie évolutionnaire au sens de l'école de Santa Barbara, dont le programme de recherche fut habilement résumé par la formule historique « nos crânes modernes abritent un esprit de l'âge de pierre » (Cosmides & Tooby, 2007). La nature stéréotypée des opérations effectuées par les modules cognitifs (au sens du modularisme massif) permet en effet d'en attendre une loi d'activation très régulière depuis leur émergence dans le passé évolutionnaire humain, et donc une individuation par leur fonction écologique ancestrale. L'hypothèse

l'évolution humaine.

2. Ne prend que certaines classes d'information ou d'entrée, où l'entrée (a) peut être externe ou interne, (b) peut être extraite activement de l'environnement ou reçue passivement de l'environnement, et (c) spécifie à l'organisme les caractéristiques particulières de son environnement.

3. Transforme cette information en sortie par une procédure (par exemple, une règle de décision) dans laquelle la sortie (a) régule l'activité physiologique, fournit des informations à d'autres mécanismes psychologiques ou produit une action manifeste et (b) résout un problème adaptatif particulier.» (D. M. Buss, 1995)

«Des modules [fonctionnellement spécialisés] ont évolué par un processus de descendance avec modification, en raison des effets qu'ils avaient sur la condition physique des organismes. Ainsi, la sélection naturelle, agissant sur les systèmes de développement qui construisent les modules au cours du développement, façonne les caractéristiques de conception des modules. Ce qui importe, d'un point de vue fonctionnel, c'est la façon dont les modules traitent l'information dans le but de réguler le comportement, car c'est ce qui a un impact sur la valeur adaptative.» (H. C. Barrett & Kurzban, 2006)

«Le cerveau doit être composé d'une grande collection de circuits, avec différents circuits spécialisés dans la résolution de différents problèmes. On peut considérer chacun de ces circuits spécialisés comme un mini-ordinateur dédié à la résolution d'un problème. Ces mini-ordinateurs spécialisés sont parfois appelés modules.» (Cosmides & Tooby, 2007)

«nos esprits sont constitués d'un grand nombre de circuits spécialisés sur le plan fonctionnel. Par exemple, nous avons des circuits neuronaux dont la conception est spécialisée pour la vision. Tout ce qu'ils font, c'est nous aider à voir. D'autres circuits neuronaux sont conçus pour l'audition. Ils ne font que détecter les changements de pression atmosphérique et en extraire des informations. Ils ne participent pas à la vision, aux vomissements, à la vanité, à la vengeance ou à quoi que ce soit d'autre. D'autres circuits neuronaux encore sont spécialisés dans l'attraction sexuelle, c'est-à-dire qu'ils régissent ce que vous trouvez sexuellement

de la modularité massive permet donc une explication de traits comportementaux complexes tels que la coopération conditionnelle (Cosmides, 1989; Cosmides & Tooby, 1992), la jalousie (D. M. Buss, 1995, 2005), ou le contenu des invariants culturels (Baumard & Boyer, 2013; Boyer, 2003; Sperber, 1994; Sperber & Hirschfeld, 2004) par appel à des régularités comportementales mesurées en laboratoire, dans la mesure où ces régularités peuvent être associés à des problèmes adaptatifs ancestraux et donc attribuées à l'activité d'un mécanisme cognitif modulaire.

Si elle se trouvait vérifiée, l'hypothèse du modularisme massif permettrait donc de motiver la démarche d'explication cognitiviste dans un cadre évolutionniste et naturaliste. En effet, un mécanisme cognitif modulaire produit par définition une réponse stéréotypée dans des conditions spécifiques, et se prête donc parfaitement à une caractérisation dans le cadre de la démarche expérimentale stimulus-réponse de la psychologie cognitiviste. De plus, sa loi d'activation contemporaine est individuée par la fonction écologique pour laquelle il a été sélectionné dans le passé évolutionnaire humain. De ce fait, sa fonction écologique ancestrale peut être inférée en vertu d'une démarche de rétro-ingénierie spéculative visant à identifier sa fonction ancestrale ², et la dite fonction en explique (ou du

excitant, ce que vous considérez comme beau, les personnes que vous aimeriez fréquenter, etc. Nous avons tous ces circuits neuronaux spécialisés parce que le même mécanisme est rarement capable de résoudre des problèmes adaptatifs différents.» (Cosmides & Tooby, 2007)

«La sortie d'un mécanisme psychologique évolué est orientée vers la solution d'un problème adaptatif spécifique.» (D. Buss, 2015).

- 2 «Une prémisses centrale de la psychologie évolutionniste est que la principale façon non arbitraire d'identifier, de décrire et de comprendre les mécanismes psychologiques est d'articuler leurs fonctions - les problèmes adaptatifs spécifiques pour lesquels ils ont été conçus par la sélection.» (D. M. Buss, 1995)

«Ceux qui étudient les espèces dans une perspective adaptationniste adoptent la démarche d'un ingénieur. En discutant du sonar chez les chauves-souris, par exemple, Dawkins procède comme suit : "...Je vais commencer en posant un problème auquel la machine vivante est confrontée, puis j'envisagerai des solutions possibles au problème qu'un ingénieur sensé pourrait envisager ; j'arriverai finalement à la solution que la nature a effectivement adoptée" (1986, pp. 21-22). Les ingénieurs déterminent les problèmes qu'ils veulent résoudre, puis ils conçoivent des machines capables de résoudre ces problèmes de manière

moins en individue) le comportement. La validité d'une telle démarche est cependant conditionnée par l'existence d'une relation d'*homologie verticale forte* entre le trait contemporain étudié et sa forme ancestrale (Smith, 2020), c'est à dire la conservation de sa fonction au cours de son histoire évolutive. Une telle propriété n'est pas vérifiée dans le cas général, puisque des traits préexistants peuvent être récupérés pour de nouvelles fonctions (Gould & Vrba, 1982), ce qui fut par exemple le cas dans la récupération de systèmes cognitifs préexistants pour la lecture (Heyes, 2019). Dans le cas général, la caractérisation fonctionnelle d'un comportement contemporain ne permet donc pas d'identifier sa fonction écologique ancestrale, et sa fonction ancestrale ne permet pas d'identifier ses mécanismes contemporains. De ce fait, le caractère massivement modulaire de l'architecture cognitive humaine doit être établi en amont pour motiver la démarche d'explication modulariste.

L'argument généralement invoqué pour défendre le modularisme massif se réfère à l'approche modulariste en biologie évolutive et développementale ³. Cette dernière part du constat que le processus

efficace. Les biologistes évolutionnistes déterminent les problèmes d'adaptation qu'une espèce donnée a rencontrés au cours de son évolution, puis se demandent : "Que ferait une machine capable de résoudre ces problèmes ? à quoi ressemblerait une machine capable de bien résoudre ces problèmes dans des conditions ancestrales ? à quoi ressemblerait-elle ?" » (Cosmides & Tooby, 2007)

- 3 « La perspective évolutionniste est l'une de ces sources précieuses, et je ne vois pas pourquoi nous devrions nous en priver. En particulier, elle place la question de la modularité dans un cadre plus large approprié. Pour citer un biologiste théorique, "le fait que le phénotype morphologique puisse être décomposé en unités organisationnelles de base, les homologues de l'anatomie comparée, a... été expliqué en termes de modularité. . . . La signification biologique de ces unités semi-autonomes est leur rôle possible en tant que 'blocs de construction' adaptatifs " (Wagner, 1995). En psychologie, cela suggère que les deux notions de module mental et d'adaptation psychologique (au sens biologique), bien que n'étant certainement pas synonymes ou coextensifs, sont néanmoins susceptibles d'être étroitement liés. Les mécanismes mentaux autonomes qui sont également des cas très plausibles d'adaptations évoluées - la reconnaissance du visage ou la lecture de l'esprit, par exemple - sont des exemples parfaits de modularité plausible. » (Sperber, 2001)

« Le point important pour notre propos, cependant, est que l'organisation modulaire est une condition préalable à l'évolutivité, ou du moins une solution extrêmement courante à celle-ci (Wagner et Altenberg, 1996 ; Raff et Raff,

d'adaptation par sélection naturelle dépend de la possibilité pour une population de maintenir une variabilité génotypique et phénotypique, et donc (sur le temps long) de générer des traits phénotypiques héritable et adaptatifs. De ce fait, le phénotype doit donc être assez robuste vis-à-vis des variations génotypiques pour permettre l'émergence de nouveaux traits par mutation ou recombinaison génétique sans pour autant menacer le développement d'un phénotype viable (A. Wagner, 2008, 2012). Les organismes tendent à s'organiser en modules développementaux et fonctionnels autonomes, des «briques» architecturales dont l'expression peut être modulée en bloc par une mutation individuelle mais dont le développement ou les gènes associés demeurent mutuellement indépendants (G. P. Wagner, 1995, 1996; G. P. Wagner & Altenberg, 1996; G. P. Wagner & Zhang, 2011). Par exemple, l'altération d'un certain gène chez la mouche drosophile permet d'inhiber le développement des yeux ou au contraire de provoquer le développement d'yeux entièrement fonctionnels à des endroits inattendus, fournissant donc au «module développemental» que constitue l'œil un signal de contrôle sans pour autant affecter sa structure. La modularité des yeux permet donc à leur structure physiologique d'évoluer sans nuire à la robustesse de leur développement ou à celle de pattes fonctionnelles, et vice-versa. Quoique les raisons précises de son émergence demeure ambiguës (Pigliucci, 2008; G. P. Wagner et al., 2007), la modularité constitue en effet une propriété universelle des systèmes évolutionnaires, motivant donc le caractère modulaire de l'architecture cognitive humaine.

Toutefois, rien dans la modularité évolutionnaire ne suggère une individuation fonctionnelle au sens postulé par le modularisme massif. La robustesse des modules développementaux vis-à-vis des mutations garantit bien une certaine continuité de leur structure développementale au cours du temps évolutionnaire. Cette continuité structurelle peut se traduire par une forte similarité de certains traits hérités d'un ancêtre commun par différentes espèces malgré la divergence de leur lignées évolutionnaire, y compris dans le cas où ces traits adoptent une fonction différente (comme l'aile de la

2000 ; Wimsatt et Schank, 2004). [...] Puisque seule une organisation modulaire peut permettre cela, la question pour nous est de savoir si oui ou non il est approprié de penser à l'esprit comme à un système biologique, soumis aux mêmes exigences d'évolutivité que tout autre système de ce type. » (Carruthers, 2006)

chauve souris et le bras humain) (G. P. Wagner, 1996). Dans le cas où la fonction d'un trait est maintenue au cours du temps évolutionnaire, elle favorise d'autant plus la conservation de leur structure et de leur fonctionnement, et permet donc à certains organes de remplir un rôle analogue dans de larges taxons (comme par exemple le foie, qui participe au filtrage des toxines et à la biosynthèse chez l'ensemble des vertébrés). La modularité architecturale implique toutefois par définition de fortes contraintes développementales définissant quels phénotypes sont accessibles par variation du génotype, limitant le processus de sélection naturelle à un rôle de moteur dans l'exploration d'un espace défini (Pigliucci, 2010). La structure de trait biologiques (*a fortiori* si ils sont modulaires) se prête donc peu à une individuation par leur fonction ancestrale, et encore moins à une explication par une démarche de rétro-ingénierie basé sur la métaphore d'un « design » par la sélection naturelle (Boudry & Pigliucci, 2013; Gould & Lewontin, 1979).

Même si un module au sens biologique était individué par sa fonction ancestrale, la notion de « fonction » telle qu'elle s'applique dans ce contexte ne traduit pas une régularité comportementale. Le modularisme massif conçoit explicitement les modules cognitifs comme de systèmes computationnels spécialisés dans l'exécution de tâches spécifiques, définies par un ensemble d'opérations effectuées sur une catégorie particulière d'input ⁴. Cette spécialisation computationnelle est généralement motivée par

4 «Notre position est donc que des mécanismes fonctionnellement spécialisés avec des entrées informationnelles formellement définissables sont caractéristiques de la cognition humaine (et non humaine) et que ces caractéristiques devraient être identifiées comme les propriétés de signal de la "modularité". [...] Ce qui importe, d'un point de vue fonctionnel, c'est la manière dont les modules traitent l'information dans le but de réguler le comportement, car c'est ce qui a un impact sur la valeur adaptative. En tant que résultat direct et indissociable de ce processus évolutif de spécialisation, les modules deviendront spécifiques à un domaine : comme ils traitent l'information de manière spécialisée, ils auront des critères d'entrée spécifiques. Seules les informations de certains types ou formats pourront être traitées par un système spécialisé.» (H. C. Barrett & Kurzban, 2006)

« Nous avons tous ces circuits neuronaux spécialisés parce que le même mécanisme est rarement capable de résoudre différents problèmes adaptatifs. [...]

l'efficience supposée d'une architecture cognitive composée de modules fonctionnellement spécialisés, en raison du coût que constituerait l'exploration exhaustive de l'espace des inférences possibles par un système computationnel généraliste (H. C. Barrett & Kurzban, 2006; Pinker, 1999). Hors, nous savons aujourd'hui que l'inférence bayésienne constitue un mécanisme inférentiel généraliste permettant l'apprentissage de concepts complexes à moindre coût (Tenenbaum et al., 2011; Ullman & Tenenbaum, 2020), et ce via un processus de synchronisation neuronale ne nécessitant pas l'exploration ni même la représentation explicite de l'espace des inférences possibles (Badcock et al., 2019; Hipólito, 2022). La « fonction » d'un module cognitif au sens biologique renvoie donc moins à la réalisation d'une tâche computationnelle particulière qu'à une opération effectuée de manière intégrée dans (et au service de) une architecture hiérarchique complexe (Bechtel, 2009a) ⁵. De ce fait, aucun module biologique ne peut par lui-même produire un comportement (ou une régularité comportementale) en particulier, et rien ne permet *a priori* de l'individuer par une loi d'activation stimulus-réponse spécifique exprimée à l'échelle d'un individu, dont la mesure *in vitro* pourrait expliquer l'expression *in eco*.

L'explication des comportements par appel à des «modules cognitifs» individuels fonctionnellement s'appuie donc sur une confusion entre la fonction du module au sens architectural, sa fonction au sens évolutionnaire, et les comportements spécifiques qu'il produit (ou, du moins, participe à produire). Les causes de cette confusion sont à rechercher dans les engagements méthodologiques du modularisme massif H. C. Barrett &

Par conséquent, le cerveau doit être composé d'une grande collection de circuits, avec différents circuits spécialisés pour résoudre différents problèmes. On peut considérer chacun de ces circuits spécialisés comme un mini-ordinateur qui se consacre à la résolution d'un seul problème. De tels mini-ordinateurs spécialisés sont parfois appelés modules. Dans un sens, on peut donc considérer le cerveau comme une collection de mini-ordinateurs spécialisés, une collection de modules. » (Cosmides & Tooby, 2007)

- 5 Par exemple, les ailes d'une chauve-souris lui permettent de transformer de l'énergie musculaire en énergie potentielle pour voler, mais cette activité engage toute sa musculature, son système cardio-respiratoire, ses organes perceptuels et son cerveau.

Kurzban (2006), dont le but est précisément de mettre en lumière ces derniers, estime que ces derniers sont «logiquement impliqué[s] par une perspective computationnelle, laquelle [postule l'existence de] mécanismes avec des entrées et des opérations formellement définissables». Les auteurs se réfèrent sans ambiguïté à une perspective computationnelle au sens cognitiviste classique du terme, dont nous avons exposé en première partie l'engagement à étudier l'activité cognitive comme un processus algorithmique. Par définition, elle conçoit les processus cognitifs comme des systèmes input-output au fonctionnement régulier, dont l'activation est conditionnée à des classe particulières de stimuli et peut produire des comportements stéréotypés (ou du moins influencer le comportement vers des expressions stéréotypées). Ces processus sont par construction individué[s] par leur fonctionnement (Piccinini, 2010), et ce dernier peut aussi bien être caractérisé en laboratoire que formé par la sélection naturel. De ce fait, le fonctionnalisme computationnel couplé à une approche évolutionniste de la cognition suffit en effet à établir l'individuation fonctionnelle des mécanismes cognitifs, et donc à motiver le programme modulariste. L'appel modulariste à la fonction ancestrale des comportements s'appuie donc par construction sur une démarche cognitiviste assimilant mécanisme cognitif et effet psychologique, et ne peut servir à justifier cette dernière.

De fait, la notion de «module cognitif» sert généralement d'intermédiaire à une stratégie explicative s'appuyant sur l'extrapolation des effets psychologiques mesurés *in vitro* pour expliquer les comportements *in eco*. Il permet donc de revendiquer un ancrage méthodologique évolutionniste et mécaniste au programme cognitiviste, sans pour autant se contraindre aux engagements méthodologiques associés aux dites approches. Le modularisme massif déforme les concepts invoqués de sorte à s'abstraire d'une étude intégrée de la structure, de la fonction, du développement et de la phylogénie des traits visés (caractéristique de la démarche d'explication évolutionnaire Bateson & Laland (2013)), et *a fortiori* d'une modélisation causale explicite de l'architecture cognitive qui sous-tend les comportements humains. Puisque les mécanismes cognitifs modulaires sont tenus pour être individué[s] par les comportements qu'ils ont été sélectionnés pour produire, leur existence suffit à motiver que les résultats de la psychologie cognitive constituent des propriétés naturelles de la cognition humaine, et non des expressions contextuelles de la structure causale sous-jacente. L'individuation des effets psychologiques induite par le modularisme massif permet donc de motiver une démarche réductionniste d'explication des faits sociaux par les faits psychologiques, antagoniste à une démarche proprement naturaliste

basée sur une intégration explicite des différentes échelles de comportement qui sous-tendent les activités humaines via une modélisation causale explicite. Nous allons à présent nous intéresser à l'extension de ces démarches en sciences sociales en comparant deux programmes d'explication des comportements humains, basés respectivement sur une démarche réductionniste et une démarche intégrative, que propose la discipline de l'évolution culturelle.

3 – Naturalisation des comportements sociaux : intégration et réduction en évolution culturelle

L'évolution culturelle est une discipline récente proposant une démarche de « naturalisation » des sciences sociales visant à intégrer leur méthodologie à celle des sciences évolutionnaires. Elle se définit avant tout par une démarche d'explication des dynamiques socio-culturelles humaines par l'héritabilité de traits culturels via l'apprentissage social, et réunit dans ce cadre conceptuel une grande diversité thématique comme méthodologique (Mesoudi, 2017; Mesoudi et al., 2006). Le fait que les traits désignés comme culturels soient (par définition (Heyes, 2020b)) issus de l'apprentissage social, et donc dépendant de l'état d'une population à l'instant précédent, n'implique en soi qu'une continuité historique des changements culturels que personne ne conteste et qui n'est que très peu restrictivement méthodologiquement. La spécificité de l'évolution culturelle comme discipline réside dans l'appel explicite à des modèles de dynamiques de populations hérité de la synthèse évolutionnaire moderne (voir eg Laland et al. (2015) pour une définition), lesquels étudient l'évolution dans une population donnée de la distribution de traits transmis selon des règles particulières entre individus biologiques (Lewens, 2015, Chapter 2). Ces modèles supposent donc par construction l'existence de répliqueurs individués – c'est à dire des entités capables de se transmettre sans modification notable entre individus biologiques, comme des gènes – condition qui ne semble pas être vérifiée pour l'évolution culturelle (Chvaja, 2020; Lewens, 2015, Chapter 2). Les entités tenues comme étant soumises à réplique sont donc typiquement les traits culturels eux-mêmes, généralement conçus de représentations cognitives ou de comportements développés lors d'interactions sociales

directes ou de l'engagement avec un environnement socialement altéré (Lewens, 2015, Chapter 3).

Cette description traduit avant tout la Gene-Culture Coevolutionary Theory (GCCT), selon laquelle l'évolution culturelle se caractérise avant tout par un processus d'adaptation par la sélection de traits phénotypique écologiquement appropriés, lequel s'appuie sur l'interaction entre héritabilité génétique et culturelle (Boyd & Richerson, 1988; Henrich & McElreath, 2013; P. J. Richerson & Boyd, 2008). Cette approche s'intéresse avant tout aux mécanismes qui sous-tendent la sélection de comportements écologiquement appropriés, et tend donc à ignorer les processus cognitifs qui sous-tendent l'acquisition des dits comportements. Certains auteurs critiquent donc cette démarche, et proposent en réponse la Cultural Attraction Theory (CAT) selon laquelle les processus cognitifs de reconstruction de représentations cognitives lors de l'apprentissage social affectent la stabilité et le changement du contenu de tokens culturels (matériels ou cognitifs) au sein d'une population (Boyer, 1998; Scott-Phillips et al., 2018; Sperber, 1996; Sperber & Hirschfeld, 2004). Elle tient de ce fait le processus de sélection comme tout au plus un cas particulier d'une dynamique plus générale d'attraction des tokens culturels vers des formes plus transmissibles socialement (Claidière, Scott-Phillips, et al., 2014), laquelle affecte aussi bien l'art (Morin, 2012; Morin & Acerbi, 2016), les pratiques médicales (Miton et al., 2015; Miton & Mercier, 2015), ou les signes linguistiques (Claidière, Smith, et al., 2014; Morin, 2018). De ce fait, tout en restant explicitement attachée à une approche populationnelle et évolutionniste des changements culturels, elle s'intéresse avant tout à la manière dont les contraintes psychologiques humaines forment les contenus culturels.

Nous allons à présent illustrer la divergence méthodologique entre CAT et GCCT par leurs traitements respectifs de l'évolution des comportements religieux, c'est à dire à l'écologie complexe de pratiques rituelles et de croyances surnaturelles qui permée de manière quasi-universelle dans les sociétés humaines, et se sont vraisemblablement développées en parallèle des autres traits définissant les humains modernes (Boyer & Bergstrom, 2008; Sterelny, 2018). Un large accord existe aujourd'hui sur le rôle critique des processus cognitifs qui sous-tendent l'attribution d'agentivité et l'intégration

de normes sociale dans le développement de ces comportements (Atran & Ginges, 2012; Atran & Henrich, 2010; Atran & Norenzayan, 2004; J. L. Barrett & Zahl, 2013; Baumard et al., 2015; Baumard & Boyer, 2013; Baumard & Chevallier, 2015; Boyer, 2003; Boyer & Baumard, 2016; Boyer & Bergstrom, 2008; Boyer & Liénard, 2006; Purzycki, 2018; Purzycki & Sosis, 2009; Sterelny, 2018; Whitehouse & Lanman, 2014). La CAT et la GCCT divergent toutefois dans le rôle explicatif précis qui leur est accordé. Si la CAT estime que ces processus traduisent l'activité de mécanismes cognitifs évolués dont les comportements religieux constituent un sous-produit (J. L. Barrett, 2000; Boyer, 2003; Boyer & Baumard, 2016), les approches associées à la GCCT considèrent au contraire que les comportements religieux s'expliquent avant tout par un processus d'adaptation culturelle exploitant les dits processus pour catalyser la coordination sociale (Atran & Henrich, 2010; Norenzayan et al., 2016; Purzycki & Sosis, 2009; Sterelny, 2018). De ce fait, bien que la CAT revendique une prise en compte précise des processus cognitifs sous-tendant l'évolution culturelle, sa démarche concrète tend à expliquer les comportements *in eco* par appel à des effets expérimentaux mesurables *in vitro*.

Cette approche est particulièrement transparente dans l'études des religions dites "moralisantes", un ensemble de systèmes religieux institutionnalisés définissant une doctrine morale explicite qui ont émergé de manière indépendante dans différentes aires socio-culturelles au cours du premier millénaire avant J.-C. (Taylor, 2012). Certains chercheurs rattachés à la CAT estiment que les religions moralisantes constituent un simple sous-produit du développement de stratégies évolutionnaires orientées à plus long terme (incluant par exemple la réflexivité, l'investissement dans des relations monogames, ou la confiance envers autrui) en raison de l'enrichissement des sociétés archaïques durant la période axiale (Baumard et al., 2015; Baumard & Chevallier, 2015). Cette thèse est essentiellement appuyé par une étude quantitative montrant le rôle causal de proxy d'affluence et de densité de population dans l'émergence des religions moralisantes (Baumard & Boyer, 2013). Cette étude échoue toutefois à individuer l'affluence comme variable explicative centrale, dans la mesure où les trois variables qu'il identifie comme prédictives (affluence, densité de population, taille de la capitale)

sont des proxys évidents de la complexité socio-politique (telle que définie dans Turchin et al. (2018)), et donc de nombreux facteurs institutionnels pouvant aussi bien traduire la pression de sélection pour un système de normes prosociales à grande échelle que l'opportunité pour un clergé préexistant de s'intégrer dans le pouvoir politique. Surtout, l'étude ne mesure aucunement la rétroaction des religions moralisantes sur l'affluence ou la complexité socio-politique, et ne peut donc absolument contredire une explication adaptationniste de l'émergence des religions moralisantes.

Ces dernières présentent pourtant toutes les marques d'une adaptation culturelle résultant d'un processus de sélection de groupe (P. Richerson et al., 2016; D. S. Wilson et al., 2008), médié par l'héritabilité culturelle des comportements religieux à l'échelle des groupes co-religioneux. Premièrement, les religions moralisantes sont caractérisées non seulement par leur contenu doctrinal mais surtout par un clergé professionnel qui en diffuse la doctrine à grande échelle via un mode rituel approprié à cet usage (doctrinal au sens de Whitehouse & Lanman (2014)). Deuxièmement, leurs doctrines encouragent systématiquement la prosocialité au-delà des communautés familiales et ethniques ainsi qu'une croyance institutionnalisée en une punition surnaturelle des comportements immoraux, favorisant l'application des normes définies par l'institution religieuse et plus généralement la coopération entre coreligioneux (Norenzayan et al., 2016; Purzycki, Henrich, et al., 2018; Purzycki, Ross, et al., 2018). Troisièmement, cette capacité à catalyser la coopération s'est historiquement traduite par une extension agressive de ces religions, issue de la diffusion au cours du temps des formes doctrinales et institutionnelles les plus aptes à maintenir la cohésion sociale des groupes humains (Turchin et al., 2019; D. Wilson, 2010). L'approche adaptationniste de l'évolution des religions moralisantes intègre donc non seulement le rôle de l'écologie et de la psychologie humaine dans les comportements religieux, mais aussi le rôle des croyances et des institutions religieuses dans la psychologie et l'écologie humaine. Par exemple, une étude récente suggère un rôle central de l'Église Catholique Romaine dans la désintégration des institutions familiales en Europe occidentale, et indirectement dans le développement de la psychologie individualiste occidentale (Schulz et al., 2019).

De fait, la GCCT reconnaît donc la capacité de l'évolution culturelle à affecter en profondeur les comportements humains. Son postulat méthodologique central, l'existence d'un processus d'adaptation dirigé par la sélection naturelle des variantes culturelles les plus appropriés, n'est en rien menacé par l'absence de répliqueurs culturels (Acerbi & Mesoudi, 2015; Godfrey-Smith, 2012) - au contraire de ce que peuvent suggérer des auteurs liés à la CAT (Claidière, Scott-Phillips, et al., 2014; Scott-Phillips et al., 2018). Elle explique certes le dit processus par des traits fonctionnels (ou comportementaux) caractérisables en conditions expérimentales, typiquement des biais dans la transmission de traits culturels portant sur leur contenu, leur contexte, ou leur fréquence (Boyd & Richerson, 1988, Chapter 5), lesquels seraient issus d'une sélection ancestrale en vertu de leur rôle dans l'adaptation culturelle (Boyd et al., 2011; Boyd & Richerson, 1988; Henrich & McElreath, 2013; P. J. Richerson & Boyd, 2008). Toutefois, cette démarche d'explication cognitiviste est remise dans le contexte d'une dynamique co-évolutionnaire entre gène et culture, et permet donc une étude systématique de la rétroaction des dynamiques culturelles sur les comportements humains, qu'il s'agisse des traits psychologiques expliquant l'évolution culturelle (Heyes, 2012, 2020a; Muthukrishna et al., 2021) ou des pratiques et institutions formant les sociétés humaines (Henrich & Muthukrishna, 2021; Shilton et al., 2020). De ce fait, l'approche de la GCCT constitue plus un schéma explicatif des dynamiques d'adaptation culturelle qu'une tentative de réduire les comportements humains à un ensemble de traits comportementaux caractérisables en laboratoire (Lewens, 2015; Smaldino, 2017; voir aussi Piccinini & Craver, 2011).

Cette démarche intégrative est à contraster avec la démarche de naturalisation des sciences sociales dont se revendiquent les chercheurs associés à la CAT (Scott-Phillips et al., 2018; Sperber, 1996). Son fondateur historique, Dan Sperber, est également le fondateur de la théorie du modularisme massif (Sperber, 2001), auquel il fait explicitement appel dans la définition de la démarche de la CAT (Sperber, 1996, p. 113; Sperber & Hirschfeld, 2004). Certains chercheurs associés ne souscrivent certes pas au modularisme massif *per se* (Boyer, 1998), et des écrits postérieurs ont explicitement souligné le rôle de facteurs environnementaux dans l'attraction

culturelle (Scott-Phillips et al., 2018; Sperber, 2011; Sperber et al., 2011). La CAT vise toutefois ouvertement à expliquer l'évolution culturelle par l'activité d'un ensemble de processus mentaux domaine-spécifiques produisant des relations input-output régulières ⁶, lesquelles sont supposés êtres prédéfinies par l'histoire évolutionnaire de l'espèce. Elle traduit donc directement une démarche d'explication cognitiviste en sciences sociales, appelant à expliquer les dynamiques socio-culturelles s'exprimant en conditions écologiques par les régularités comportementales mesurés par la démarche expérimentale stimulus-réponse de la psychologie cognitive. Dan Sperber va même plus loin en soutenant que l'explication scientifique des phénomènes sociaux doit s'ancrer dans la description des processus cognitifs sous-jacents ou renoncer à toute prétention causale ⁷, et que les sciences

6 «Je propose une naturalisation du domaine des sciences sociales non pas sur le modèle de la psychologie mais sur celui de l'épidémiologie. Autrement dit, je veux faire valoir que les phénomènes sociaux sont des modèles de phénomènes psychologiques et écologiques à l'échelle d'une population. Ce projet de naturalisation du social est rendu possible par la naturalisation du mental qui est en cours dans les sciences cognitives.

Laissez-moi vous expliquer comment. Ce qui rend un processus cognitif cognitif est qu'il a pour fonction d'assurer une relation de contenu entre son entrée et sa sortie. Dans le cas de la perception, l'entrée est un stimulus, la sortie est une représentation mentale, et le processus de perception vise à garantir que le contenu de la représentation mentale soit une véritable identification du stimulus d'entrée. [...] Ce que les sciences cognitives sont en train de faire, c'est d'expliquer comment les processus matériels peuvent garantir de manière fiable de telles relations entre les contenus de représentation (comme dans l'inférence ou la mémoire) ou entre les contenus et les états d'affaires qu'ils représentent (comme dans la perception et l'action).» (Sperber, 2011)

7 «L'hypothèse, c'est qu'avec une description matérielle, naturelle, de ce qui se passe à l'intérieur des organismes individuels et dans leur environnement commun, on aura tout ce dont on a besoin pour rendre compte du social. Sinon, tant pis pour le programme naturaliste... Quand on dit que tout est là, cela ne veut pas dire qu'on renonce à expliquer ce que cherchent à expliquer les sciences sociales (le sens, les institutions, les normes, etc.), mais on va essayer de le faire à partir [de chaînes causales cognitives] et de leur articulation.»

«La plupart des approches mécanistes, que leur individualisme soit seulement méthodologique ou aussi métaphysique, montrent peu d'intérêt à doter les sciences sociales d'une ontologie naturaliste, c'est-à-dire continue avec celle des sciences naturelles. [...] Bien sûr, des descriptions de type causal sans contrainte

sociales devraient donc importer le langage et les entités supposées par le modularisme massif (Sperber, 2011). Bien qu'il se défende de soutenir un programme de réduction théorique ⁸, il n'en appelle pas moins à une réduction scientifique des phénomènes socio-culturels aux phénomènes cognitifs.

Cette discussion devrait exposer l'ambivalence profonde des notions de "naturalisation" ainsi que de "démarche cognitive". Tout les défenseurs d'un programme de naturalisation des sciences sociales sont d'accord pour voir l'architecture cognitive comme la base des comportements sociaux (ce en quoi ils sont rejoints par les défenseurs du *statu quo* en sciences sociales (Ogien, 2011)). Un désaccord profond existe toutefois sur la nécessité d'expliquer les dynamiques socio-culturelles par appel aux propriétés universelles de nature humaine, ou d'expliquer au contraire la construction de la cognition humaine par appel aux dynamiques socio-culturelles. Nous avons exposé ici que le premier programme tend à s'appuyer sur une démarche d'explication cognitive au sens fort, laquelle suppose l'individuation et l'universalité des différents effets psychologiques mesurables par les moyens de la psychologie cognitive. Elle s'écarte donc de fait des standards de l'explication naturaliste, au sens défini ici d'une démarche explicative visant à cartographier les structures causales dans le monde, et rejoint plutôt une démarche positiviste de réduction des comportements humains aux lois mesurables en laboratoire. Au contraire, la seconde enrichit la démarche naturaliste des neurosciences cognitives en exposant comment se développent les mécanismes cognitifs qui sous-tendent la cognition humaine eux-mêmes (Heyes, 2018, 2020a). Il apparaît donc

ontologique peuvent suffire à l'objectif poursuivi, mais on ne vise alors pas vraiment des explications causales scientifiques des phénomènes sociaux. En d'autres termes, se limiter à une ontologie naturaliste est une condition favorable - je suis tenté de dire, nécessaire - pour parvenir à des affirmations causales solides.»

(Sperber et al., 2011)

- 8 «La façon dont les sciences sociales naturalisées renoncent à l'autonomie ontologique garantit leur autonomie théorique. ontologique garantit leur autonomie théorique. En d'autres termes, je plaide pour une réduction ontologique sans réduction théorique.» (Sperber, 2011)

qu'une approche naturaliste rigoureuse de la cognition humaine ne puisse s'abstraire de son caractère historique, construit, et situé, et diverge ainsi radicalement d'une démarche d'explication cognitiviste des comportements. Nous allons à présent nous intéresser à la nature physique des processus que nous désignons sous le nom de "cognition", et nous allons chercher à établir pourquoi la mesure des régularité comportementales exprimées *in vitro* ne peut suffir à expliquer les dynamiques cognitives s'exprimant *in eco*.

4 – L'architecture physique de la cognition : autopoïèse et rupture de symétrie multi-échelle

5 – Restaurer l’unité des sciences cognitives ? Le FEP comme paradigme intégratif des dynamiques cognitives

Conclusion

Remerciements

Je souhaite remercier ici Benoît de Courson et le regretté Martin Fortier pour le soutien qu'ils ont apporté au lancement de ce projet, ainsi que Iona Brenac et Yseult Héjja-Brichard pour leurs relectures et commentaires sur le document présent. Je remercie également le collectif Kairos pour le cadre stimulant qu'il a offert au travail de synthèse scientifique et épistémologique que j'ai présenté ici.

Bibliographie

Acerbi, A., & Mesoudi, A. (2015). If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. *Biology & Philosophy*, 30(4), 481–503.
<https://doi.org/10.1007/s10539-015-9490-2>

Andler, D. (2010). Naturalism and the Scientific Status of the Social Sciences. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association* (pp. 1–12). Springer Netherlands.
https://doi.org/10.1007/978-90-481-3263-8_1

Atran, S., & Ginges, J. (2012). Religious and Sacred Imperatives in Human Conflict. *Science*, 336(6083), 855–857.
<https://doi.org/10.1126/science.1216902>

Atran, S., & Henrich, J. (2010). The Evolution of Religion: How Cognitive By-Products, Adaptive Learning Heuristics, Ritual Displays, and Group Competition Generate Deep Commitments to Prosocial Religions. *Biological Theory*, 5(1), 18–30. https://doi.org/10.1162/BIOT_a_00018

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences*, 27(6), 713–730.
<https://doi.org/10.1017/S0140525X04000172>

- Badcock, P. B., Friston, K. J., & Ramstead, M. J. D. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews*, 31, 104–121. <https://doi.org/10.1016/j.plrev.2018.10.002>
- Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113(3), 628–647. <https://doi.org/10.1037/0033-295X.113.3.628>
- Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, 4(1), 29–34. [https://doi.org/10.1016/S1364-6613\(99\)01419-9](https://doi.org/10.1016/S1364-6613(99)01419-9)
- Barrett, J. L., & Zahl, B. P. (2013). Cognition, evolution, and religion. In *APA handbook of psychology, religion, and spirituality (Vol 1): Context, theory, and research* (pp. 221–237). American Psychological Association. <https://doi.org/10.1037/14045-012>
- Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology & Evolution*, 28(12), 712–718. <https://doi.org/10.1016/j.tree.2013.09.013>
- Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in Cognitive Sciences*, 17(6), 272–280. <https://doi.org/10.1016/j.tics.2013.04.003>
- Baumard, N., & Chevallier, C. (2015). *The nature and dynamics of world religions: A life-history approach*.
- Baumard, N., Hyafil, A., Morris, I., & Boyer, P. (2015). Increased Affluence Explains the Emergence of Ascetic Wisdoms and Moralizing Religions.

Current Biology, 25(1), 10–15.

<https://doi.org/10.1016/j.cub.2014.10.063>

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Taylor & Francis.

Bechtel, W. (2009a). Explanation: Mechanism, modularity, and situated cognition. *The Cambridge Handbook of Situated Cognition*, 155–170.

Bechtel, W. (2009b). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, 22(5), 543–564.

<https://doi.org/10.1080/09515080903238948>

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421–441. <https://doi.org/10.1016/j.shpsc.2005.03.010>

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, 193(5), 1509–1534. <https://doi.org/10.1007/s11229-015-0783-4>

Boudry, M., & Pigliucci, M. (2013). The mismeasure of machine: Synthetic biology and the trouble with engineering metaphors. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4, Part B), 660–668. <https://doi.org/10.1016/j.shpsc.2013.05.013>

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. University of Chicago Press.

- Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(Supplement 2), 10918–10925. <https://doi.org/10.1073/pnas.1100290108>
- Boyer, P. (1998). Cognitive Tracks of Cultural Inheritance: How Evolved Intuitive Ontology Governs Cultural Transmission. *American Anthropologist*, 100(4), 876–889. <https://doi.org/10.1525/aa.1998.100.4.876>
- Boyer, P. (2003). Religious thought and behaviour as by-products of brain function. *Trends in Cognitive Sciences*, 7(3), 119–124. [https://doi.org/10.1016/S1364-6613\(03\)00031-7](https://doi.org/10.1016/S1364-6613(03)00031-7)
- Boyer, P., & Baumard, N. (2016). The Diversity of Religious Systems Across History: An Evolutionary Cognitive Approach. In *The Oxford Handbook of Evolutionary Psychology and Religion*.
- Boyer, P., & Bergstrom, B. (2008). Evolutionary Perspectives on Religion. *Annual Review of Anthropology*, 37(1), 111–130. <https://doi.org/10.1146/annurev.anthro.37.081407.085201>
- Boyer, P., & Liénard, P. (2006). Why ritualized behavior? Precaution Systems and action parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences*, 29(6), 595–613. <https://doi.org/10.1017/S0140525X06009332>
- Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00599>

- Buss, D. (2015). *Evolutionary Psychology: The New Science of the Mind*. Psychology Press.
- Buss, D. M. (1995). Evolutionary Psychology: A New Paradigm for Psychological Science. *Psychological Inquiry*, 6(1), 1–30.
https://doi.org/10.1207/s15327965pli0601_1
- Buss, D. M. (2005). *The dangerous passion: Why jealousy is a necessary as love and sex*. Odile Jacob.
- Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.
- Chvaja, R. (2020). Why Did Memetics Fail? Comparative Case Study1. *Perspectives on Science*, 28(4), 542–570.
https://doi.org/10.1162/posc_a_00350
- Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642), 20130368.
<https://doi.org/10.1098/rstb.2013.0368>
- Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797), 20141541. <https://doi.org/10.1098/rspb.2014.1541>
- Cognitive revolution. (2021). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Cognitive_revolution&oldid=1057303159

- Colombo, M., Hartmann, S., & van Iersel, R. (2015). Models, Mechanisms, and Coherence. *The British Journal for the Philosophy of Science*, 66(1), 181–212. <https://doi.org/10.1093/bjps/axt043>
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187–276. [https://doi.org/10.1016/0010-0277\(89\)90023-1](https://doi.org/10.1016/0010-0277(89)90023-1)
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, 163, 163–228.
- Cosmides, L., & Tooby, J. (2007). *Evolutionary Psychology: A primer*.
- Craver, C. F. (2014). The ontic conception of scientific explanation. *Explanation in the Biological and Historical Sciences. Dordrecht, Netherlands: Springer*, 27–52.
- Cummins, R. C. (2000). ‘How Does It Work’ Versus ‘What Are the Laws?’: Two Conceptions of Psychological Explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and Cognition*, 117-145. MIT Press.
- Dohmatob, E., Dumas, G., & Bzdok, D. (2020). Dark control: The default mode network as a reinforcement learning agent. *Human Brain Mapping*, n/a(n/a). <https://doi.org/10.1002/hbm.25019>
- Dumas, G., Kelso, S., & Nadel, J. (2014). Tackling the social cognition paradox through multi-scale approaches. *Frontiers in Psychology*, 5, 882. <https://doi.org/10.3389/fpsyg.2014.00882>

- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-Brain Synchronization during Social Interaction. *PLOS ONE*, 5(8), e12166. <https://doi.org/10.1371/journal.pone.0012166>
- Felin, T., Felin, M., Krueger, J. I., & Koenderink, J. (2019). On Surprise-Hacking. *Perception*, 48(2), 109–114. <https://doi.org/10.1177/0301006618822217>
- Felin, T., Koenderink, J., & Krueger, J. I. (2017). Rationality, perception, and the all-seeing eye. *Psychonomic Bulletin & Review*, 24(4), 1040–1059. <https://doi.org/10.3758/s13423-016-1198-z>
- Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.
- Girard, B., Tabareau, N., Pham, Q. C., Berthoz, A., & Slotine, J.-J. (2008). Where neuroscience and dynamic system theory meet autonomous robotics: A contracting basal ganglia model for action selection. *Neural Networks*, 21(4), 628–641. <https://doi.org/10.1016/j.neunet.2008.03.009>
- Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2160–2170. <https://doi.org/10.1098/rstb.2012.0118>
- Gould, S. J., & Lewontin, R. C. (1979). *The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme*.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—A Missing Term in the Science of Form. *Paleobiology*, 8(1), 4–15. <https://doi.org/10.1017/S0094837300004310>

- Graham, G. (2019). Behaviorism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/spr2019/entries/behaviorism/>
- Henrich, J., & McElreath, R. (2013). *The Evolution of Cultural Evolution*.
- Henrich, J., & Muthukrishna, M. (2021). The Origins and Psychology of Human Cooperation. *Annual Review of Psychology*, 72(1), 207–240.
<https://doi.org/10.1146/annurev-psych-081920-042106>
- Heyes, C. (2012). Grist and mills: On the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2181–2191.
<https://doi.org/10.1098/rstb.2012.0120>
- Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University Press.
- Heyes, C. (2019). Précis of Cognitive Gadgets: The Cultural Evolution of Thinking. *Behavioral and Brain Sciences*, 42.
<https://doi.org/10.1017/S0140525X18002145>
- Heyes, C. (2020a). Psychological Mechanisms Forged by Cultural Evolution. *Current Directions in Psychological Science*, 29(4), 399–404.
<https://doi.org/10.1177/0963721420917736>
- Heyes, C. (2020b). *Culture primer* [Preprint]. PsyArXiv.
<https://doi.org/10.31234/osf.io/v8ms3>

- Hipólito, I. (2022). Cognition Without Neural Representation: Dynamics of a Complex System. *Frontiers in Psychology*, 12, 5472. <https://doi.org/10.3389/fpsyg.2021.643276>
- Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 93(5), 1449–1475. <https://doi.org/10.1257/000282803322655392>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Kaufmann, L., & Cordonier, L. (2011). Vers un naturalisme social. À la croisée des sciences sociales et des sciences cognitives. *Sociologies*. <http://journals.openedition.org/sociologies/3595>
- Ladyman, J., Ross, D., Collier, J., Collier, with D. S. and J., Spurrett, D., Spurrett, D., & Collier, J. G. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Clarendon Press.
- Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., & Odling-Smee, J. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proc. R. Soc. B*, 282(1813), 20151019. <https://doi.org/10.1098/rspb.2015.1019>
- Lewens, T. (2015). *Cultural Evolution: Conceptual Challenges*. OUP Oxford.
- Longo, G., & Montévil, M. (2013). Extended criticality, phase spaces and enablement in biology. *Chaos, Solitons & Fractals*, 55, 64–79. <https://doi.org/10.1016/j.chaos.2013.03.008>

- Longo, G., Montévil, M., & Kauffman, S. (2012). No entailing laws, but enablement in the evolution of the biosphere. *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*, 1379–1392. <https://doi.org/10.1145/2330784.2330946>
- Longo, G., Montévil, M. R., & Pocheville, A. (2012). From Bottom-Up Approaches to Levels of Organization and Extended Critical Transitions. *Frontiers in Physiology*, 3. <https://doi.org/10.3389/fphys.2012.00232>
- Macy conferences. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Macy_conferences&oldid=1058627126
- Marr, D., Poggio, T., & Brenner, S. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 204(1156), 301–328. <https://doi.org/10.1098/rspb.1979.0029>
- Mesoudi, A. (2017). Pursuing Darwin's curious parallel: Prospects for a science of cultural evolution. *Proceedings of the National Academy of Sciences*, 114(30), 7853–7860. <https://doi.org/10.1073/pnas.1620741114>
- Mesoudi, A., Whiten, A., & Kevin, L. (2006). *Towards a unified science of cultural evolution*.
- Mitchell, S. D. (2003). *Biological Complexity and Integrative Pluralism*. Cambridge University Press.

- Miton, H., Claidière, N., & Mercier, H. (2015). *Universal cognitive mechanisms explain the cultural success of bloodletting*.
- Miton, H., & Mercier, H. (2015). *Cognitive Obstacles to Pro-Vaccination Beliefs*.
- Molenaar, P. C. M., & Nesselroade, J. R. (2015). Systems methods for developmental research. In *Handbook of child psychology and developmental science: Theory and method, Vol. 1, 7th ed* (pp. 652–682). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9781118963418.childpsy117>
- Morin, O. (2012). *How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution*.
- Morin, O. (2018). Spontaneous Emergence of Legibility in Writing Systems: The Case of Orientation Anisotropy. *Cognitive Science*, 42(2), 664–677. <https://doi.org/10.1111/cogs.12550>
- Morin, O., & Acerbi, A. (2016). *Birth of the cool: A two-centuries decline in emotional expression in Anglophone fiction*.
- Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a Historical Science. *Annual Review of Psychology*, 72(1), 717–749.
<https://doi.org/10.1146/annurev-psych-082820-111436>
- Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences*, 39.
<https://doi.org/10.1017/S0140525X14001356>

- Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, 3(8), 782–791. <https://doi.org/10.1038/s41562-019-0626-2>
- Ogien, A. (2011). Les sciences cognitives ne sont pas des sciences humaines. Une réponse à « Vers un naturalisme social » de Laurence Kaufman et Laurent Cordonier. *Sociologies*. <http://journals.openedition.org/sociologies/3635>
- Papineau, D. (2020). Naturalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/naturalism/>
- Piccinini, G. (2010). The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism. *Philosophy and Phenomenological Research*, 81(2), 269–311. <https://doi.org/10.1111/j.1933-1592.2010.00356.x>
- Piccinini, G., & Craver, C. (2011). Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. *Synthese*, 183(3), 283–311. <https://doi.org/10.1007/s11229-011-9898-4>
- Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics*, 9(1), 75. <https://doi.org/10.1038/nrg2278>
- Pigliucci, M. (2010). Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal*

Society B: Biological Sciences, 365(1540), 557–566.

<https://doi.org/10.1098/rstb.2009.0241>

Pinker, S. (1999). *How the Mind Works*. W.W. Norton.

Purzycki, B. G. (2018). Institutions, natural selection(s), and religion.

Religion, Brain & Behavior, 0(0), 1–8.

<https://doi.org/10.1080/2153599X.2018.1513860>

Purzycki, B. G., Henrich, J., Apicella, C., Atkinson, Q. D., Baimel, A.,

Cohen, E., McNamara, R. A., Willard, A. K., Xygalatas, D., &

Norenzayan, A. (2018). The evolution of religion and morality: A synthesis of ethnographic and experimental evidence from eight societies.

Religion, Brain & Behavior, 8(2), 101–132.

<https://doi.org/10.1080/2153599X.2016.1267027>

Purzycki, B. G., Ross, C. T., Apicella, C., Atkinson, Q. D., Cohen, E.,

McNamara, R. A., Willard, A. K., Xygalatas, D., Norenzayan, A., &

Henrich, J. (2018). Material security, life history, and moralistic religions: A cross-cultural examination. *PLoS ONE*, 13(3).

<https://doi.org/10.1371/journal.pone.0193856>

Purzycki, B. G., & Sosis, R. (2009). The Religious System as Adaptive:

Cognitive Flexibility, Public Displays, and Acceptance. In E. Volland &

W. Schiefenhövel (Eds.), *The Biological Evolution of Religious Mind and Behavior* (pp. 243–256). Springer. [https://doi.org/10.1007/978-3-](https://doi.org/10.1007/978-3-642-00128-4_17)

642-00128-4_17

- Ramstead, M. J. D. (2015). Naturalizing what? Varieties of naturalism and transcendental phenomenology. *Phenomenology and the Cognitive Sciences*, 14(4), 929–971. <https://doi.org/10.1007/s11097-014-9385-8>
- Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M., & Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, 39. <https://doi.org/10.1017/S0140525X1400106X>
- Richerson, P. J., & Boyd, R. (2008). *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.
- Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, 366(6466). <https://doi.org/10.1126/science.aau5141>
- Scott-Phillips, T., Blanke, S., & Heintz, C. (2018). Four misunderstandings about cultural attraction. *Evolutionary Anthropology: Issues, News, and Reviews*, 27(4), 162–173. <https://doi.org/10.1002/evan.21716>
- Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human Social Evolution: Self-Domestication or Self-Control? *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00134>
- Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In *Computational Social Psychology*. Routledge.
- Smith, S. E. (2020). Is Evolutionary Psychology Possible? *Biological Theory*, 15(1), 39–49. <https://doi.org/10.1007/s13752-019-00336-4>

- Speelman, C., & McGann, M. (2013). How Mean is the Mean? *Frontiers in Psychology*, 0. <https://doi.org/10.3389/fpsyg.2013.00451>
- Speelman, C. P., & McGann, M. (2016). Editorial: Challenges to Mean-Based Analysis in Psychology: The Contrast Between Individual People and General Science. *Frontiers in Psychology*, 0. <https://doi.org/10.3389/fpsyg.2016.01234>
- Speelman, C. P., & McGann, M. (2020). Statements About the Pervasiveness of Behavior Require Data About the Pervasiveness of Behavior. *Frontiers in Psychology*, 11, 3117. <https://doi.org/10.3389/fpsyg.2020.594675>
- Sperber, D. (1994). The epidemiology of representations. *Mapping the Mind: Domain Specificity in Cognition and Culture*, 39.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach*.
- Sperber, D. (2001). In defense of massive modularity. *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, 7, 47–57.
- Sperber, D. (2011). A naturalistic ontology for mechanistic explanations in the social sciences. *Analytical Sociology and Social Mechanisms*, 64–77.
- Sperber, D., Coubray, A., & Schmitt, Y. (2011). Entretien avec Dan Sperber. Naturalisme, sciences cognitives et religion. *ThéoRèmes. Penser le religieux*. <https://doi.org/10.4000/theoremes.153>
- Sperber, D., & Hirschfeld, L. A. (2004). *The cognitive foundations of cultural stability and diversity*.

- Sterelny, K. (2018). Religion re-explained. *Religion, Brain & Behavior*, 8(4), 406–425. <https://doi.org/10.1080/2153599X.2017.1323779>
- Taylor, C. (2012). 2 What Was the Axial Revolution? In *The Axial Age and Its Consequences* (pp. 30–46). Harvard University Press. <https://www.degruyter.com/document/doi/10.4159/harvard.9780674067400.c2/html>
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, 331(6022), 1279–1285. <https://doi.org/10.1126/science.1192788>
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Turchin, P., Currie, T. E., Whitehouse, H., François, P., Feeney, K., Mullins, D., Hoyer, D., Collins, C., Grohmann, S., Savage, P., Mendel-Gleason, G., Turner, E., Dupeyron, A., Cioni, E., Reddish, J., Levine, J., Jordan, G., Brandl, E., Williams, A., ... Spencer, C. (2018). Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proceedings of the National Academy of Sciences*, 115(2), E144–E151. <https://doi.org/10.1073/pnas.1708800115>
- Turchin, P., Whitehouse, H., François, P., Hoyer, D., Nugent, S., Larson, J., Covey, A., Altaweel, M., Peregrine, P., Carballo, D., Feinman, G., Wallace, V., Bol, P. K., Korotayev, A., Kradin, N., Anderson, E., Savage, P. E., Cioni, E., Levine, J., ... Squitieri, A. (2019). *Explaining the Rise of Moralizing Religions: A test of competing hypotheses using*

the Seshat Databank [Preprint]. SocArXiv.

<https://doi.org/10.31235/osf.io/2v59j>

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian Models of Conceptual Development: Learning as Building Models of the World. *Annual Review of Developmental Psychology*, 2(1), 533–558.

<https://doi.org/10.1146/annurev-devpsych-121318-084833>

van Riel, R., & Van Gulick, R. (2019). Scientific Reduction. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019).

Metaphysics Research Lab, Stanford University.

<https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/>

Wagner, A. (2008). Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630),

91–100. <https://doi.org/10.1098/rspb.2007.1137>

Wagner, A. (2012). The role of robustness in phenotypic adaptation and innovation. *Proceedings of the Royal Society B: Biological Sciences*,

279(1732), 1249–1258. <https://doi.org/10.1098/rspb.2011.2293>

Wagner, G. P. (1995). Adaptation and the modular design of organisms. In F. Morán, A. Moreno, J. J. Merelo, & P. Chacón (Eds.), *Advances in Artificial Life* (pp. 315–328). Springer. [https://doi.org/10.1007/3-540-](https://doi.org/10.1007/3-540-59496-5_308)

[59496-5_308](https://doi.org/10.1007/3-540-59496-5_308)

Wagner, G. P. (1996). Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist*, 36(1), 36–43.

<https://doi.org/10.1093/icb/36.1.36>

- Wagner, G. P., & Altenberg, L. (1996). Perspective: Complex Adaptations and the Evolution of Evolvability. *Evolution*, 50(3), 967–976.
<https://doi.org/10.1111/j.1558-5646.1996.tb02339.x>
- Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, 8(12), 921–931.
<https://doi.org/10.1038/nrg2267>
- Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: The evolvability of complex organisms. *Nature Reviews Genetics*, 12(3), 204–213.
<https://doi.org/10.1038/nrg2949>
- Whitehouse, H., & Lanman, J. A. (2014). The Ties That Bind Us: Ritual, Fusion, and Identification. *Current Anthropology*, 55(6), 674–695.
<https://doi.org/10.1086/678698>
- Wilson, D. (2010). Darwin’s Cathedral. In *Darwin’s Cathedral*. University of Chicago Press.
<https://www.degruyter.com/document/doi/10.7208/9780226901374/html>
- Wilson, D. S., Van Vugt, M., & O’Gorman, R. (2008). Multilevel Selection Theory and Major Evolutionary Transitions: Implications for Psychological Science. *Current Directions in Psychological Science*, 17(1), 6–9. <https://doi.org/10.1111/j.1467-8721.2008.00538.x>
- Wright, C., & Bechtel, W. (2007). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science*

(pp. 31–79). North-Holland. <https://doi.org/10.1016/B978-044451540-7/50019-0>