



HAL
open science

La neuroéconomie en question : débats et controverses

Daniel Serra

► **To cite this version:**

| Daniel Serra. La neuroéconomie en question : débats et controverses. 2019. halshs-02160911

HAL Id: halshs-02160911

<https://shs.hal.science/halshs-02160911>

Preprint submitted on 20 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La neuroéconomie en question : debats et controverses

Daniel Serra



CEE-M Working Paper 2019-13

LA NEUROÉCONOMIE EN QUESTION : DÉBATS ET CONTROVERSES

Daniel Serra

CEE-M, Univ. Montpellier, CNRS, INRA, Montpellier SupAgro, Montpellier, France

2 juin 2019

Résumé

Comme toute discipline émergente, la neuroéconomie soulève bon nombre de questions donnant lieu à débats et controverses. L'article en rend compte en les regroupant autour de trois questions majeures qui portent sur le statut des connaissances à ce jour produites : Des connaissances certaines ? Des connaissances scientifiquement justifiées ? Des connaissances utiles à l'économiste ? Sans prétendre à l'exhaustivité, ce cadre d'analyse permet d'apprécier les lignes de force des principales critiques adressées à la neuroéconomie en même temps que la nature des parades susceptibles d'être déployées.

Mots-clés : Neuroéconomie, Economie comportementale, Economie expérimentale, Méthodologie économique, Epistémologie scientifique

Abstract

As any nascent discipline, neuroeconomics raises many questions. The article reports on the main debates and controversies that neuroeconomics has aroused in the literature. Three major issues relating to the status of knowledge produced are addressed. Are results reliable? Are they scientifically legitimate? What is their usefulness for an economist? This framework allows to assess the main criticisms addressed to neuroeconomics and the nature of the parades which can be deployed.

Keywords : Neuroeconomics, Behavioral economics, Experimental economics, Economic methodology, Scientific epistemology

Classification JEL :

D87, B41, C9

* Corresponding Author: Daniel Serra, CEE-M, Univ. Montpellier, CNRS, INRA, Montpellier SupAgro, Montpellier, France ; Faculté d'économie, Site Richter, Avenue de la mer – Raymond Dugrand, 34960, Montpellier, CS : 79606 ; e-mail: daniel.serra@univ-montp.fr.

† L'auteur remercie Sacha Bourgeois-Gironde, Philippe Mongin, Pierre Livet, Nicolas Vallois pour des commentaires ou des conversations très utiles sur les sujets abordés dans cet article.

Le cerveau depuis toujours fascine les hommes. La « neuroéconomie », née aux États-Unis au tournant des années 2000, est la toute dernière discipline scientifique à placer ce mystérieux organe au cœur de ses investigations. Indiscutablement, le décollage de l'économie expérimentale dans les années 1980 a favorisé l'émergence des premières études de neuroéconomie en offrant un ensemble de *designs* expérimentaux bien codifiés pour traiter des problèmes de décision économique individuelle ou en interaction. On s'accorde cependant à reconnaître que cette discipline émerge, à la jonction entre l'économie, la psychologie et la biologie, s'appuie essentiellement sur la révolution des neurosciences cognitives une dizaine d'année auparavant et la mise à disposition de tout un éventail d'outils d'investigation sophistiqués pour mener les expériences, avec au premier plan l'imagerie par résonance magnétique fonctionnelle (IRMf) grâce à laquelle les corrélats métaboliques de l'activité neuronale peuvent être mesurés de manière non invasive.

Si la neuroéconomie fait l'objet aujourd'hui d'un véritable engouement dans les milieux scientifiques, à l'instar de l'économie expérimentale en ses débuts, elle reste malgré tout controversée chez les économistes. La neuroéconomie suscite tout à la fois fascination et scepticisme ; d'autant que les débats demeurent intenses voire acerbes entre spécialistes. L'objet de cet article est de rendre compte de ces débats et controverses en les ordonnant autour de trois questions majeures. Une première série de critiques vise la fiabilité des résultats. Cela concerne à la fois le nombre fort limité de données sur lesquelles ils reposent souvent, la fiabilité des outils mobilisés, les difficultés que l'on rencontre dans l'interprétation de ces données neuronales, voire la logique qui préside à certaines expériences (des connaissances certaines ?) (section 1). Une seconde catégorie de critique, de nature plus philosophique, s'interroge sur la pertinence de données non comportementales (données cognitives, physiologiques, neuronales) pour tester des modèles de choix économiques : une question épineuse de légitimité scientifique qui concerne aussi certains travaux menés en économie comportementale (des connaissances scientifiquement justifiées ?) (section 2). Au-delà des questionnements relatifs à la fiabilité des résultats des expériences neuroéconomiques et à la légitimité de la démarche, certains économistes interrogent enfin l'utilité pour l'économie des enseignements tirés de ces expériences neuroéconomiques, au regard principalement des résultats déjà bien établis par l'économie comportementale (des connaissances utiles ?) (section 3).

1. LA FIABILITÉ DES RÉSULTATS ?

Parmi les multiples critiques adressées à la neuroéconomie, nombreuses sont celles qui visent des questions de pratiques ou de « méthodes » (en les distinguant des questions de « méthodologie » ou d'épistémologie qui seront abordées dans la section 2). Elles soulèvent le problème de la fiabilité des résultats, de la confiance que l'on peut leur accorder. L'économiste Glenn Harrison est certainement l'auteur dont la critique est à cet égard la plus virulente. Les problèmes évoqués concernent autant la production des données (elles seraient peu nombreuses et peu fiables) que leur interprétation (corrélation ou causalité ?). Certaines critiques s'adressent plus particulièrement aux études d'imagerie cérébrale : elles visent notamment la fiabilité de l'outil comme les conditions de validité de la logique qui préside aux résultats (les « inférences inverses »).

1.1. La production des données

Un certain nombre d'auteurs ont critiqué la tendance de la neuroéconomie à tirer des conclusions quelque peu hâtives sur la base de données trop peu nombreuses (Rubinstein, 2006, 2008 ; Harrison, 2008a) et peu

fiables (Harrison, 2008a, 2008b). Il est vrai que cette critique relative à la faiblesse des échantillons n'est pas sans fondement. Malgré tout, elle perd progressivement de sa pertinence avec la réalisation d'expériences mobilisant des échantillons plus conséquents en relation avec l'élargissement de la gamme des outils employés et, surtout, l'apparition des expériences de terrain dans des contextes naturels. Quant aux « simplifications statistiques » qui grèveraient la fiabilité des résultats, elles visent essentiellement la technique de l'imagerie cérébrale dans les neurosciences en général.

Le faible nombre de données

Si l'on compte le cerveau comme unité dans l'analyse, les premières expériences portaient effectivement sur un nombre très faible de données : autour d'une dizaine. Quelques études, plus récemment, concernent des effectifs de sujets un peu plus étoffés : par exemple, les expériences menées par Lorhens *et al.* (2007) et Smith *et al.* (2014) portent sur 54 et 35 sujets respectivement. Mais c'est vrai que la plupart des expériences employant l'imagerie cérébrale portent encore aujourd'hui, essentiellement pour des raisons budgétaires, sur des échantillons de taille modeste (pas plus d'une vingtaine bien souvent)¹. Pour autant, on ne doit pas perdre de vue que l'unité dans l'analyse n'est pas le cerveau : c'est l'activation par unité de temps d'une certaine zone localisée dans le cerveau. De sorte que, au final, on dispose dans chaque étude d'un ensemble relativement conséquent de données issues d'un nombre limité de cerveaux, chacun contribuant à la production de plusieurs observations en chaque point du temps. Ce qui affaiblit quelque peu la critique : Harrison (2008a) du reste le reconnaît (voir également Harrison et Ross, 2010).

Dans un ordre d'idées différent, il convient de mentionner le développement d'une technologie d'imagerie cérébrale (*hyper-scanning technology*) qui permet de scanner *simultanément* deux (ou plus de deux) sujets en interaction en utilisant des connections en réseau entre deux (ou plusieurs) IRMf séparés (Montague *et al.*, 2002). Cette technologie innovante, en permettant de collecter au moins deux fois plus de données sans que cela prenne plus de temps, ouvre en particulier des perspectives nouvelles pour la neuroéconomie « sociale »². Mais l'accroissement du nombre des données traitées devrait passer aussi par le développement des expériences de terrain.

Les expériences de terrain : une voie à développer

La neuroéconomie *in situ* est rendue possible grâce aux apports de la neuroendocrinologie et de la psychophysiologie qui donnent accès à cette batterie extensive d'instruments peu onéreux et susceptibles d'être employés sur le terrain (administration de substances, réponse électrodermale, oculométrie, etc.). Elle constitue une réponse peu onéreuse à la question de la taille des échantillons de sujets. Elle apporte en outre un élément de réponse à la question de la « validité externe » des expériences de neuroéconomie (*i.e.* la généralisation des résultats en dehors du laboratoire), une question dont on ne s'est pas vraiment préoccupé jusqu'à aujourd'hui alors qu'elle a fait l'objet de réflexions poussées en économie expérimentale (voir notamment Guala, 2005 ; Heukelom, 2009 ; Bardsley *et al.*, 2010 ; Serra 2012b).

L'une des premières expériences de ce type a été menée en 2007 par John Coates (un trader de *Wall Street* reconverti en neurophysiologiste) dans une salle d'échange d'une banque située à Londres (Coates et Herbert, 2008). Analysant les taux de certaines d'hormones chez des traders au cours de leur journée de travail, il détermina sur une période de plusieurs jours comment ces taux fluctuaient en liaison avec leur performance financière et la volatilité du marché. Les hormones présentes dans le sang constituent le moyen par lequel le cerveau communique avec les différents organes ; elles fournissent à cet égard une

¹ En économie expérimentale, les expériences de première génération portaient également sur des tailles d'échantillon aussi réduites. Un aperçu historique de l'économie expérimentale est proposé par Serra (2012a, 2015) (voir également Cot et Ferey, 2016).

² Un terme introduit par Fehr et Camerer (2007) pour évoquer les travaux de neuroéconomie consacrés aux situations d'interactions sociales.

information sur la manière dont le cerveau réagit à un certain contexte. Lors de cette expérience, les « sujets » se trouvaient bien dans un environnement de travail « naturel », ayant la capacité de gagner ou de perdre réellement de l'argent³. Lo et Repin (2002) ont également étudié les réactions émotionnelles de traders en fonction des fluctuations du marché à l'aide de certains outils de la psychophysologie (conductance de la peau, pulsations cardiaques, etc.). Cette expérience pilote menée avec seulement 10 professionnels a été prolongée par Lo, Repin et Steenbarger (2005) avec un échantillon de sujets plus large (80 sujets volontaires) et en recourant à une méthode complémentaire pour évaluer leurs réponses émotionnelles (réponses à des questionnaires avant et après des périodes d'entraînement d'une durée de cinq semaines auprès d'un organisme de formation de traders).

D'une manière générale, ces expériences de terrain portent sur des échantillons de sujets bien plus conséquents que les expériences en laboratoire. Ainsi, par exemple, l'expérience fameuse sur la confiance conduite par Kosfeld *et al.* (2005) porte sur un échantillon de 194 sujets (voir plus loin dans cette section pour une discussion plus large sur l'influence des hormones dans les comportements).

Des simplifications statistiques discutables ?

À supposer que les observations sur lesquelles s'appuient les travaux expérimentaux soient en nombre suffisant, le problème des données ne disparaît pas pour autant – en tout cas dans les expériences d'imagerie cérébrale. Car les méthodes employées dans le traitement statistique de ces images neuronales sous forme de séries chronologiques ainsi que les procédures d'agrégation utilisées pour parvenir à un « cerveau moyen » soulèvent de nombreuses interrogations. Harrison (2008b), en particulier, pointe ces « simplifications statistiques » qui, selon lui, grèvent largement la fiabilité des résultats des expériences. Il prétend en particulier que la normalisation des données aboutit à une surestimation des seuils de significativité, à une minimisation des variations individuelles et par suite à l'identification plus facile des aires neuronales activées au cours des tâches accomplies par les sujets. Harrison (2008a), dont le langage est très imagé, parle de véritable « usine à gaz » (lorsqu'il évoque l'analyse statistique des données spatio-temporelles recueillies sur un cerveau, de « boîte de conserve infectée de vers » s'agissant des inférences nécessaires à l'agrégation des données de plusieurs cerveaux, qualifiant d'« économétrie à la MacGyver » (du nom du fameux héros d'une série télévisée américaine célèbre pour son ingéniosité face à tout problème pratique) le problème plus large de la modélisation statistique des données neuronales opérée dans les expériences de neuroéconomie ! Il reproche en outre aux auteurs de ces expériences de ne jamais rendre publiques leurs données neuronales. Une pratique héritée de la tradition en neurosciences qui s'oppose à cet égard à celle des économistes expérimentalistes qui transmettent toujours leurs données à la demande. C'est la question de la répliquabilité qui est bien sûr ici visée. On peut trouver des critiques voisines – sur le fond et non dans la forme ! – dans Platt et Huettel (2008), Huettel (2010), Kable (2011) ou Van Rooij et Van Orden (2011)⁴.

Bien que cette critique ne soit pas spécifique à la neuroéconomie – elle s'adresse aux expériences qui emploient la neuro-imagerie en général –, elle ne saurait être sous-estimée. Elle contribue en tout cas à militer en faveur de la conclusion générale que l'on peut tirer de l'examen de ces critiques visant la fiabilité

³ L'étude révéla que le taux de testostérone (une hormone qui intervient non seulement en matière de sexualité mais aussi dans les comportements impulsifs et à risque) et celui de cortisol (une hormone dont le niveau s'accroît avec le stress) étaient fortement influencés par la volatilité du marché. Un enseignement important de cette étude est ainsi que ces deux hormones – l'une favorisant un comportement impulsif et l'autre accompagnant le stress – soufflent « le chaud et le froid » en ce sens qu'elles doivent rester à l'intérieur d'un certain intervalle de concentration pour qu'un trader se comporte efficacement sur le marché.

⁴ La publication de ce dernier article a malgré tout suscité des échanges plutôt « musclés » entre Colin Camerer et Alec Smith d'un côté (soutenus par une vingtaine d'économistes et neuroscientifiques travaillant à l'interface des deux disciplines) et Mariëk Van Rooij et John Holden (un collaborateur de longue date de Van Orden, décédé entre temps) de l'autre côté (Camerer et Smith ; Van Rooij et Holden, 2013).

des résultats des expériences de neuroéconomie, à savoir encourager le recours à une pluralité d'outils d'investigation⁵.

1.2. L'interprétation des données

D'autres critiques visent l'interprétation des données que génère tel ou tel outil d'investigation (Fudenberg, 2006; Poldrack, 2006 ; Harrison, 2008a, 2008b ; Harrison et Ross, 2010 ; van Rooij et van Orden, 2011). La neuroéconomie, là encore, se heurte à des difficultés qui ne lui sont pas propres mais qui concernent les neurosciences cognitives en général ; chacun des instruments qui intervient dans les expériences en neurosciences possède des avantages mais aussi des limites (Cravel, 2007). Ainsi, si les outils d'imagerie cérébrale servent à *localiser* les aires cérébrales dont l'activation coïncide avec la prise de décision, il est clair qu'elles ne suffisent pas à établir l'implication *causale* de ces aires. Identifier certaines zones neuronales est une chose ; comprendre ce qui s'y passe et comment elles interviennent dans le fonctionnement du cerveau est ce qui importe véritablement. On dispose heureusement aujourd'hui en neuroéconomie d'une batterie d'outils d'intervention qui complète les outils de localisation et permet d'établir de réelles causalités et non de simples corrélations. Mais Fehr (2009, p. 226) mentionne aussi la possibilité de générer de véritables causalités par le seul choix de *designs* expérimentaux appropriés. C'est ainsi qu'on peut se servir de l'activité neuronale observée lors d'un « traitement » particulier pour prédire un comportement lors d'un autre « traitement » (prévision « hors traitement »). Il cite l'exemple de l'expérience pionnière menée en neurosciences sociales par de Quervain *et al.* (2004) sur une variante du « jeu de la confiance », dans laquelle les auteurs se servent des différences individuelles observées dans l'activation du *noyau caudé*⁶ lorsque la punition des sujets non coopératifs n'est pas coûteuse pour ceux qui l'infligent (traitement initial) pour prédire combien d'individus vont effectivement payer pour punir lorsqu'elle est coûteuse (traitement différent).

Parmi les nombreuses difficultés soulevées par l'interprétation des images d'IRMf figure également l'interprétation du « mode par défaut » du cerveau. Car, même au repos, le cerveau manifeste une certaine activité (voir notamment Forest, 2014, et Raichle, 2015). L'imagerie cérébrale par IRMf souffre en outre d'une contrainte technique : le signal BOLD mesuré par les voxels qui composent l'image se révèle très faible. Ce qui explique en particulier que les expériences de neuroéconomie menées à l'aide de cette technologie nécessitent la répétition de la même tâche plusieurs fois de manière à retenir la *moyenne* des signaux. Par suite, l'étude des décisions interactives dans les jeux « à un coup » (*one-shot*) ne peut être menée à l'aide des protocoles expérimentaux classiques ; ils doivent être modifiés pour tenir compte de cette contrainte technique. On utilise des jeux « à plusieurs coups » (*multi-shot*) : le jeu est répété plusieurs fois « en *stranger* »⁷ (avec un tirage au sort de l'une des répétitions pour rémunérer les sujets (Houser et McCabe, 2009, p. 55). De même, toute autre cause de variation du signal, comme un mouvement de la tête du sujet dans le scanner, doit être strictement contrôlée. De façon à minimiser ce biais, la tête des sujets est souvent stabilisée à l'aide de coussins en mousse, les stimuli visuels étant projetés sur un écran à l'arrière du scanner de manière à ce que les sujets puissent les voir grâce à un miroir. Certains auteurs vont

⁵ Il existe plus généralement une littérature récente en philosophie des sciences qui mène une réflexion critique sur l'usage de l'imagerie cérébrale comme technique expérimentale de preuve dans les neurosciences cognitives – mais que l'on peut transférer sans difficulté à la neuroéconomie – avec des auteurs comme William Bechtel, Valerie Hardcastle, Stewart Matthew, Adina Rooskies et William Uttal. Voir Forest (2014), qui brosse un tableau riche et contrasté de cette littérature. On trouvera dans Poldrack *et al.* (2017) des éléments de réponse à ces questionnements et un plaidoyer pour un code de « bonnes pratiques » en la matière.

⁶ Les noyaux de la base (ou noyaux gris centraux) représentent la structure sous-corticale la plus ancienne au plan de l'évolution (tous les vertébrés, sous une forme ou une autre) possèdent cette structure cérébrale. Elle comprend un ensemble fonctionnellement diversifié de noyaux enfouis dans la profondeur des hémisphères cérébraux, derrière les lobes frontaux, comme le striatum (ou corps strié). Le noyau caudé (*caudate*) correspond à la partie la plus supérieure du striatum.

⁷ La procédure d'appariement « en *stranger* » consiste à faire en sorte que les mêmes sujets ne soient jamais appariés entre eux plusieurs fois.

jusqu'à octroyer un bonus financier supplémentaire aux sujets qui respectent scrupuleusement la condition d'immobilité de la tête (par exemple, Smith *et al.*, 2014)⁸.

Au-delà des critiques visant les limites attachées à l'outil utilisé, il est possible de formuler une critique supplémentaire qui s'attache à la conception même de certaines expériences de neuroéconomie. Ainsi, pour Harrison (2008a), bien des chercheurs dans les expériences portant sur les décisions sociales ou les choix intertemporels négligeraient dans leur interprétation des données comportementales observées l'existence possible d'artefacts ou de confusions (*confounds*). Un phénomène pourtant bien dévoilé désormais en économie expérimentale (voir notamment Bardsley *et al.*, eds., 2010, et Serra, 2012b, pour une discussion de cette question en économie expérimentale). « Il est clair », nous dit Harrison, « que nous avons un travail conceptuel à effectuer avant de mettre le scanner en marche » (2008a, p. 321). Faute de quoi, le risque est grand pour le chercheur de promouvoir l'explication qu'il a en tête en concevant l'expérience au détriment d'explications alternatives, sans réelle preuve expérimentale. Harrison prend notamment l'exemple du « jeu de la confiance » pour lequel les neuroéconomistes, sans précaution selon lui, avancent des explications discutables. Ils interprètent par exemple les comportements coopératifs observés comme une manifestation de la confiance que les investisseurs portent à leurs partenaires, alors que les motivations peuvent se révéler extrêmement variées (altruisme, goût intense pour le risque, désir d'infliger des pertes monétaires élevées à l'expérimentaliste...) (Sanfey *et al.*, 2003, par exemple). De même, ils interprètent les comportements de « punition » comme générateur d'une utilité « émotionnelle » alors que cela peut être considéré comme parfaitement rationnel de punir ceux qui se détournent d'une stratégie profitable si l'on tient compte du fait que les sujets peuvent percevoir la situation comme un jeu répété, bien que celui-ci ne le soit pas, puisque joué « en *stranger* » (De Quervain *et al.*, 2004, par exemple). La tromperie des sujets, contre laquelle pourtant les expérimentalistes en économie ont développé tout un argumentaire, est une autre cause potentielle de confusion dans l'interprétation de certaines expériences de neuroéconomie (Gallagher *et al.*, 2002 ; Rilling *et al.*, 2004 ; Knoch *et al.*, 2006a, notamment). Il est vrai toutefois que les économistes s'opposent en cela à la pratique des psychologues (voir Serra, 2012b)⁹.

Les techniques de localisation : de simples corrélations

Dans la mesure où il existe un niveau élevé d'interaction entre les diverses régions du cerveau et puisque l'activation d'une zone spécifique du cerveau n'est pas toujours parfaitement corrélée à une motivation ou un mécanisme cognitif clairement défini, il paraît difficile d'inférer un élément causal à partir d'effets en termes de processus et de fonctions neuronales. Comme le rappelaient déjà Kenning et Plassman (2005) en commentant les premières expériences de neuroéconomie, bien souvent, elles « sont seulement capables de mettre en évidence une corrélation entre l'accomplissement d'une tâche spécifique et une activité cérébrale dans certaines zones spécifiques. Mais cela ne doit pas être compris comme la preuve d'une relation causale réelle » (p. 352).

Si les expériences d'imagerie cérébrale ne permettent pas de se prononcer sur la manière dont les fonctions ou états mentaux sont implémentés dans le cerveau, il convient toutefois de signaler qu'il existe une catégorie particulière d'expériences qui présentent un avantage par rapport aux approches d'IRMf ordinaires, les expériences « basées sur des modèles » (*model-based approach*). Dans ces expériences,

⁸ En dépit des précautions prises lors des expériences par IRMf, il est encore fréquent de devoir exclure de l'analyse certains sujets du fait de mouvements de tête excessifs lors de l'acquisition de l'image (voir par exemple l'étude récente de Nagel *et al.*, 2018, qui teste l'hypothèse selon laquelle les choix en incertitude stratégique pourraient se ramener à des choix en situation risquée, dans laquelle les observations portant sur 2 des 20 sujets ont dû être exclus de l'analyse pour ce motif).

⁹ On peut trouver dans Serra (2016) une discussion de la plupart de ces expériences visées par la critique de Harrison.

plutôt que de corrélérer directement les données neuronales et les données comportementales (« approche « directe »), les corrélations sont médiatisées par un modèle computationnel explicite. Dans les cas les plus simples, la particularité de cette approche est de partir d'un modèle de calcul décrivant la transformation d'un ensemble de stimuli en un ensemble de réponses comportementales. Les opérations « internes » spécifiques nécessaires à la réalisation d'une telle transformation correspondent alors aux variables pertinentes dans l'étude de neuroéconomie, puisque ce sont ces variables qui seront éventuellement corrélées aux données neuronales. En incorporant dans les analyses d'imagerie cérébrale les paramètres comportementaux estimés dans un modèle computationnel explicite, il devient ainsi possible de savoir non seulement « où » mais « comment » une fonction particulière est implémentée dans le cerveau. Les données neuronales, en complément des paramètres comportementaux, deviennent ici une puissante source de contrainte pour les modèles de calcul neuronaux (O'Doherty, Hampton et Kim, 2007, Corrado et Doya, 2007, et Corrado *et al.*, 2009, fournissent des descriptions détaillées des expériences « basées sur des modèles » en neuroéconomie). Comme le soulignent O'Doherty, Hampton et Kim (2007), ces expériences présentent malgré tout une limite : elles restreignent les chances de détecter des résultats *inattendus*. En conséquence, « une politique raisonnable consiste probablement à employer les analyses plus conventionnelles basées sur une série d'essais en parallèle de l'approche basée sur un modèle » (p. 50). En outre, il est vrai que ces expériences ne permettent toujours pas de parvenir à des explications causales.

Au final, avec les techniques de *localisation*, l'interprétation des images neuronales reste plutôt délicate. C'est pourquoi ces études doivent s'adjoindre des techniques d'*intervention*.

Les techniques d'intervention : de réelles causalités

La majorité des difficultés liées à l'interprétation des données cérébrales disparaissent avec l'introduction de manipulations expérimentales, telles que la stimulation transcrânienne ou l'administration de substances, destinées à induire de manière exogène certaines conséquences comportementales. L'usage de ces méthodes, qui se développe rapidement depuis une dizaine d'années, permet d'établir en neuroéconomie de réelles *causalités* et non de simples corrélations, en complément des enseignements issus des études impliquant des patients atteints de lésions ou souffrant de certaines pathologies qui restent un vecteur de connaissances majeur en la matière¹⁰.

La stimulation transcrânienne

La stimulation magnétique transcrânienne répétée (SMT) et la stimulation transcrânienne en courant continu (STCC) donnent toutes deux la possibilité d'accroître (effet excitateur) ou de diminuer (effet inhibiteur) l'activité neuronale de régions spécifiques du cortex avant que les sujets ne prennent leurs décisions lors de l'exécution d'une tâche expérimentale destinée à éliciter leurs préférences. La STCC jouit d'un avantage supplémentaire : elle permet de contourner une difficulté que l'on rencontre quand on emploie des outils d'imagerie cérébrales ou la SMT en neuroéconomie sociale dans l'étude des « préférences sociales » (*i.e.* quand on suppose que les individus portent une certaine « attention aux autres » : altruisme, réciprocité bienveillante, aversion à l'inéquité...). Cette problématique oblige en effet à confronter chaque joueur à une série de jeux à un coup joués en *stranger* de manière séquentielle en privilégiant l'étude comportementale de celui qui joue en second. Ce qui pose un problème de mise en œuvre car, chaque sujet devant faire face à un nombre élevé de protagonistes, la tentation de tromper les sujets et de les confronter à des choix préfabriqués est grande. Or, en économie expérimentale, je l'ai déjà

¹⁰ On peut trouver dans Ruff et Huettel (2014) et Genon *et al.*, 2018) une évaluation des avantages et des limites des différents méthodes utilisées dans les expériences de neuroéconomie.

rappelé, il est fortement recommandé de ne pas décevoir les sujets afin d'éviter la perte de confiance dans l'expérimentaliste. La STCC, compte tenu de son faible coût d'utilisation, rend possible une intervention simultanée sur un grand nombre de sujets qui interagissent effectivement (voir par exemple Knoch *et al.*, 2008). Dans le cadre des expériences par imagerie cérébrale, il existe toutefois une autre manière de traiter cette difficulté : elle consiste à confronter chaque sujet aux choix opérés *préalablement* par des partenaires en interaction lors de jeux identiques. C'est cette stratégie expérimentale qui a été employée notamment dans l'expérience séminale menée par de Quervain *et al.* (2004). Dans cette expérience par tomographie par émission de positons (TEP), chaque sujet était confronté aux choix opérés par certains des sujets qui avaient participé préalablement à l'expérience pilote ; à l'issue de celle-ci, l'accord des sujets avait été obtenu pour que les données recueillies soient utilisées dans une expérience future, sachant qu'ils seraient rémunérés alors en fonction de leurs propres choix. Se reporter à Fehr (2009) pour une discussion plus poussée sur ces différents aspects méthodologiques.

Quelques résultats peuvent être donnés à titre illustratif. On a pu montrer, entre autres, qu'une variation induite au plan expérimental de certaines zones du cerveau (le *cortex préfrontal dorsolatéral* en général¹¹) peut modifier la disponibilité à payer des sujets pour acquérir certains biens (Camus *et al.*, 2009), les rendre moins averses au risque (Knoch *et al.*, 2006a), plus impatientes (Figner *et al.*, 2010), plus égoïstes (Knoch *et al.*, 2006b ; Baumgartner *et al.*, 2011), moins réceptifs au sentiment de culpabilité (Krajbich *et al.*, 2009), plus enclins à violer certaines normes sociales (Ruff, Ugazio et Fehr, 2013) ou encore qu'elle peut modifier leur capacité à résister à la tentation (Knoch et Fehr, 2007) ou à se construire une réputation favorable (Knoch *et al.*, 2009), voire influencer sur leur propension spontanée au mensonge (Karton et Bachmann, 2011 ; Baumgartner *et al.*, 2013).

D'un autre côté, on sait aujourd'hui que certaines manipulations pharmacologiques affectent également le cerveau et génèrent d'importants effets causaux. L'endocrino-économie, s'appuyant sur certains mécanismes physiologiques connus, est sans doute en passe de devenir l'une des orientations majeures de la neuroéconomie. Parmi les substances chimiques en jeu on trouve essentiellement des hormones (cortisol, testostérone, adrénaline...) ou des neuropeptides (ocytocine, arginine vasopressine...), et plus rarement des produits psychotropes (benzodiazépines, nicotine, amphétamines, cocaïne, héroïne...). Toutes ces substances sont supposées avoir une influence sur les comportements en agissant comme neurotransmetteur ou via leur action sur certains neurotransmetteurs du système nerveux central (dopamine, sérotonine, noradrénaline, acétylcholine...).

Les manipulations pharmacologiques

À ce jour, l'ocytocine est de loin la substance qui a fait l'objet des investigations expérimentales les plus nombreuses. Ce neuropeptide, synthétisé essentiellement par l'hypothalamus, est souvent considérée comme une « hormone féminine ». Elle intervient d'abord dans le renforcement des liens primaires entre la mère et son enfant en agissant sur l'allaitement (Insel, 2010). Mais elle assurerait ensuite, plus généralement, une fonction de socialisation plus large (Churchland, 2011). En employant le « jeu de la confiance » comme paradigme comportemental pour tester la confiance des investisseurs après inhalation nasale d'ocytocine, Kosfeld *et al.* (2005) établissent que l'ocytocine favoriserait la confiance en élevant la propension des individus à prendre des risques « sociaux » (voir aussi Baumgartner *et al.*, 2008, Mikolajczak *et al.*, 2010). L'une des propriétés de l'ocytocine est en effet d'inhiber l'activation de

¹¹ Situé sur la paroi extérieure du lobe frontal, les aires neuronales du cortex préfrontal latéral sont fortement impliquées dans les tâches cognitives comme la prise de décision cohérente, la délibération, la planification, la mémoire exécutive, etc. Dans sa partie supérieure, le cortex dorsolatéral est considéré comme la partie la plus « rationnelle » du cerveau ; il est associé essentiellement aux tâches de contrôle des décisions et de maintien de leur but.

l'amygdale¹². En revanche, l'inhalation d'ocytocine n'influencerait pas la loyauté des mandataires, bénéficiaires de la confiance. Pour expliquer cette asymétrie entre investisseurs et mandataires, les auteurs se réfèrent à une distinction entre la confiance « pure », décelable chez les investisseurs (elle ne peut naître que d'une certaine empathie), et la confiance « calculée », qui se manifeste chez les mandataires (car fonction de leur expérience au cours du jeu)¹³. Pour autant, ce neuropeptide ne serait pas sans effet sur la réciprocité. Car les mandataires, en interprétant le comportement des investisseurs comme un signal de la confiance placée en eux, seraient soumis à un accroissement de décharge d'ocytocine eux-mêmes (Zak, Kurzban et Matzner, 2005b). À noter toutefois que ce dernier résultat ne résulte pas d'une expérience consistant à *administrer* des doses de cette hormone aux sujets (*i.e.* en modifiant de manière *exogène* leur état physiologique) et à comparer les résultats avec un groupe « placebo » ; il s'agit plutôt d'une comparaison du niveau *endogène* d'ocytocine dans le sang des mêmes sujets lorsqu'ils jouent au jeu de la confiance et à un jeu standard impersonnel. Dans Zak, Kurzban et Matzner (2004), c'est un protocole similaire qui dévoile la corrélation positive qui existe entre le niveau d'ocytocine et le niveau de la confiance des investisseurs.

L'ocytocine aurait également d'autres effets dévoilés cette fois dans des expériences avec le « jeu de l'ultimatum » : elle accroîtrait la générosité (Zak, Stanton et Amadi, 2007 ; Vercoe et Zak, 2010) et réduirait l'aversion à la trahison (De Dreu, 2012a). Elle serait également corrélée avec l'expérience subjective d'empathie (Barraza et Zak, 2009). Les effets de l'inhalation nasale de ce neuropeptide ont aussi été étudiés dans le « jeu du dilemme du prisonnier » répété. S'agissant d'un jeu symétrique (les sujets interviennent à tour de rôle en premier et en second), aucune différence n'est observable entre les deux rôles sous l'effet de cette intervention. Ce qui n'est plus le cas au niveau de la réciprocité : l'ocytocine favorise les manifestations de réciprocité positive, sans influencer de manière significative les réactions de réciprocité négative (Rilling et Sanfey 2011 ; Rilling *et al.*, 2012). Les effets sur la coopération réciproque d'un autre neuropeptide, l'arginine vasopressine (AV), sont également étudiés dans ces deux expériences. L'AV est connue pour son action anxiogène (elle a tendance à causer de l'anxiété, du stress, de l'angoisse) (Heinrichs, Von Dayans et Domes, 2009). Les auteurs s'attendaient dès lors à ce que l'administration de ce neuropeptide induise au plan neuronal une activation accrue de l'amygdale, et au plan comportemental, une baisse de la coopération. Les résultats montrent au contraire que chez les sujets traités par AV le taux de réciprocité positive est plus élevé que chez les sujets traités par ocytocine ou chez ceux traités par un placebo. Les auteurs avancent une explication fondée sur le fait que l'AV agirait sur la connexion entre l'amygdale et la partie antérieure de l'*insula*¹⁴, de sorte que l'adoption d'un comportement coopératif pourrait s'interpréter comme une réponse visant à minimiser le risque de désagrément social encouru. L'influence de l'ocytocine sur l'altruisme pur a été également testée dans un « jeu de don » à une institution caritative (Barraza *et al.*, 2011) : l'ocytocine a tendance à favoriser un comportement de générosité indirecte, en ayant un impact positif sur le montant des dons monétaires octroyés par les sujets à l'institution.

Dans des synthèses portant sur de nombreux travaux expérimentaux dédiés au rôle de l'ocytocine sur les comportements, ce neuropeptide apparaît bien comme étroitement lié aux interactions sociales et au

¹² L'amygdale, constituée par un groupe de noyaux situé dans le lobe temporal médian, est considérée comme le siège des émotions négatives provoquées notamment par la peur et l'angoisse.

¹³ Schmidt et Livet (2014) proposent une distinction voisine entre « confiance-cadre » et « confiance-pari » (pp 168-171).

¹⁴ L'*insula* (ou cortex insulaire) est une partie du cortex déplacé dans la profondeur du sillon latéral, à la jonction des lobes frontaux-temporaux. Sa partie antérieure, qui entretient des connexions denses avec en particulier les centres de l'olfaction et du goût, est fortement impliquée dans le traitement des émotions, essentiellement négatives. Cette zone du cerveau agit comme un système de surveillance informant sur des situations à haut risque ou désagréables, sources potentielles de danger ou de douleur.

comportement pro-social (MacDonald et MacDonald, 2010 ; Zak, 2011 ; Riedl et Javor, 2012 ; Shamay-Tsoory *et al.*, 2013). D'une manière générale, pourtant, on doit reconnaître que la littérature récente sur la question conduit à des résultats plutôt contrastés. Ainsi, pour certains auteurs, l'ocytocine est présentée essentiellement comme un anxiolytique, présupposant que les relations sociales génèrent un climat anxigène (Evans *et al.*, 2013). Pour d'autres, elle n'influencerait de manière significative la confiance (ou réduirait la méfiance) qu'envers des proches ou des membres du groupe d'appartenance du sujet (De Dreu, 2012b). Pour d'autres encore, elle pourrait même avoir un effet inverse envers des partenaires anonymes (Pfeiffer, 2013 ; Lambert *et al.*, 2014). Globalement, même s'il semble bien que la prise d'ocytocine ait un effet chez des patients souffrant de trouble du comportement social (comme l'autisme ou les troubles d'anxiété sociale), la relation entre ocytocine et confiance chez les humains reste une question ouverte. Les preuves que l'ocytocine influence directement un comportement de confiance restent rares. Si cet effet existe, il est vraisemblablement modéré par une variété de facteurs (Nave, Camerer et McCulloch, 2017 ; Koppel *et al.*, 2017). En outre, une étude récente menée sur un jeu de dilemme social avec sanctions (Aydogan *et al.*, 2017) montre que l'inhalation intranasale d'ocytocine aurait plutôt tendance à amplifier la manifestation d'émotions sociales négatives comme la colère, en favorisant l'apparition de comportements destinés à sanctionner les *free-riders*. En définitive, il est clair que la nature de la sociabilité à laquelle l'ocytocine est étroitement associée gagnerait à être clarifiée.

La testostérone, hormone mâle par excellence, a également donné lieu à quelques expériences de manipulation pharmacologique. C'est ainsi que le jeu de la confiance a aussi été étudié « sous testostérone ». Dans ce cas, les résultats sont opposés à ceux qui sont censés résulter de l'inhalation d'ocytocine : la testostérone a tendance à réduire la propension à faire confiance, en ce sens que le niveau des « offres » des investisseurs se révèle plus faible (voir Zak *et al.*, 2009, pour une expérience avec des sujets de sexe masculin uniquement)¹⁵. Pour autant, contrairement à l'idée reçue, la testostérone ne conduit pas toujours à l'adoption de comportements agressifs. Si plusieurs études ont bien établi des *corrélations* positives entre des niveaux élevés de testostérone dans le sang et des comportements agressifs, aucune preuve expérimentale de l'influence *causale* de cette hormone dans l'agressivité n'a été fournie. C'est même un résultat contraire qui se dégage de l'expérience d'Eisenegger *et al.* (2011), menée cette fois sur un jeu de l'ultimatum, et avec des sujets de sexe féminin exclusivement (car leur niveau initial de testostérone dans le sang est *a priori* plus faible). Ces chercheurs partent de l'hypothèse selon laquelle, si la testostérone favorise l'agressivité, son administration devrait conduire à ce que le niveau des offres soit plus faible. Or, dans leur expérience les femmes qui ont reçu une dose de cette hormone (et non un placebo) ont choisi des niveaux d'offre significativement plus élevés. Comment interpréter ce résultat ? Par le fait que la peur de l'affront social qui résulterait d'un refus de l'offre l'emporte au final sur la poursuite de l'intérêt purement financier. Dans un environnement humain socialement complexe, on s'attend à ce que cela soit l'adoption d'un comportement pro-social et non celle d'un comportement agressif qui confère un certain statut social, *i.e.* une certaine valorisation aux yeux des tiers. Cette référence aux tiers (que tient implicitement l'expérimentaliste dans les expériences) est largement mobilisée dans l'ouvrage que Schmidt et Livet (2014) consacrent à l'analyse des interactions sociales. Dans les expériences menées en économie, toutefois, de nombreuses précautions sont en général prises afin de minimiser l'intervention de ces facteurs externes importées dans le laboratoire (double anonymat, répétitions des tâches en *stranger*, etc.) (voir notamment Serra, 2012b). Mais même dans ce cadre aseptisé du laboratoire, difficile d'écarter complètement ces facteurs idiosyncratiques qui imprègnent les préférences des sujets (*homegrown preferences*) (Smith, 2010). Eisenegger *et al.* (2011) fondent donc ici leur interprétation des comportements

¹⁵ Pour une revue de littérature sur la « biologie de la confiance » (effets de l'ocytocine et de la testostérone essentiellement), voir Riedl et Javor (2012).

observés en laboratoire lors de leur expérience sur la persistance de cette donnée fondamentale de l'intersubjectivité dans la vie en société. Cette intéressante expérience établit par ailleurs que la manipulation physiologique des « répondants » n'a aucun effet sur le taux d'acceptation des offres ; ce qui suggère que cette hormone n'aurait pas tendance à promouvoir un comportement altruiste¹⁶. D'autres résultats, de nature différente, ont également été établis dans des expériences administrant de la testostérone aux sujets : il semble en particulier que cette hormone encourage l'adoption d'un comportement honnête (Wibral *et al.*, 2012).

Bien d'autres substances sont administrées aux sujets dans les expériences de neuroéconomie. On a pu établir notamment que la sérotonine élève le taux de rejet des offres inéquitables dans le jeu de l'ultimatum (Crockett *et al.*, 2008) alors que l'administration de benzodiazépine le réduit (Gospic *et al.*, 2011), ou encore que la sérotonine, en façonnant le jugement moral, a tendance à accroître l'aversion à nuire à autrui (Siegel et Crockett, 2013).

On peut trouver dans Crockett et Fehr (2014) un panorama plus complet de cette littérature dans une étude qui distingue les effets de ces divers neuropeptides, hormones ou autres substances en matière de préférence pour le présent, d'aversion au risque et de préférences sociales¹⁷.

S'agissant globalement de l'emploi des manipulations pharmacologiques, Fehr (2009, p. 222) y voit une opportunité de transformation radicale de la neuroéconomie (il parle même de « révolution »). Car le faible coût d'utilisation de cette méthode expérimentale autorise la réalisation d'expériences avec des échantillons de sujets de taille comparable à celle menée en économie expérimentale. De fait, on assiste ces dernières années dans l'étude des comportements à une multiplication des expériences avec intervention pharmacologique¹⁸.

1.3. Le problème de l'« inférence inverse » (*reverse inference*)

Dans les expériences neuroéconomiques par imagerie cérébrale, les activités neuronales sont supposées associées à certaines fonctions cognitives ou certains états mentaux. Il est dès lors indispensable que la liaison entre l'activation de la zone neuronale et la fonction cognitive ou l'état mental à l'étude soit garantie. La validité interne d'une expérience type de neuroéconomie par *imagerie cérébrale*, qui repose sur une inférence des activités neuronales aux fonctions cognitives ou aux états mentaux, est conditionnée par cette hypothèse. Or cela ne va pas de soi. Les conditions de validité, sinon la légitimité, de ce raisonnement qualifié d'« inférence inverse » peuvent être discutées¹⁹.

L'indétermination de la valeur fonctionnelle de l'activité neuronale

¹⁶ Les différences comportementales et physiologiques entre sexes ont aussi été abordées par Zak *et al.* (2005a), qui établissent que la méfiance serait associée à un accroissement du taux de dihydrotestostérone (DHT) et que les hommes mais non les femmes répondraient à la méfiance par un accroissement du taux de cette hormone. La DHT est un dérivé biologiquement actif de la testostérone (testostérone « 8 octanes ») qui cause en particulier le développement des caractéristiques sexuelles secondaires des mâles durant la puberté (croissance des muscles, système pileux sur le visage, épaississement des cordes vocales...).

¹⁷ On commence en outre à avoir une idée des bases neurophysiologiques des effets comportementaux exercés par ces substances chez l'animal (essentiellement les rats) (voir Stoop, 2014, pour une synthèse des résultats).

¹⁸ Voir également Stanton (2009) pour un plaidoyer en faveur des études hormonales et plus généralement le recours au niveau *moléculaire* pour comprendre les comportements.

¹⁹ Le problème de l'« inférence inverse » n'est pas spécifique à la neuroéconomie ; il est également présent dans les expériences de psychologie cognitive qui s'appuient sur l'imagerie cérébrale pour inférer l'engagement de fonctions cognitives particulières (mémoire, attention, langage...) à partir de l'activation de régions neuronales particulières. Les premières réflexions méthodologiques consacrées à la question concernaient du reste uniquement la psychologie cognitive (D'Esposito *et al.*, 1998). Voir également Poldrack (2006), qui est le premier auteur à y avoir associé certaines expériences de neuroéconomie (Poldrack *et al.*, 2017, discutent ce problème en même temps que les autres limites qui grèvent les études de neuro-imagerie tout en suggérant en ensemble de « bonnes pratiques » en la matière).

L'accumulation des données d'imagerie cérébrale au cours des vingt dernières années forme une cartographie fonctionnelle stable du cerveau qui s'enrichit avec le temps. Dans les expériences d'imagerie cérébrale, en neuroéconomie, on se réfère alors à ces études antérieures qui établissent la correspondance entre l'activation d'une zone neuronale particulière et l'engagement d'un certain processus cognitif, pour en inférer « en sens inverse » que l'activité observée de cette même zone neuronale dans les nouvelles expériences reflète l'engagement du processus cognitif dont il est question. Harrison (2008b) qualifie ce raisonnement de « fallacieux » (p. 535), car il repose sur une inversion de la démarche courante des neurobiologistes qui consiste à localiser d'abord certaines fonctions cognitives dans des zones neuronales, puis à déduire l'engagement de ces fonctions prédéfinies de l'activation de ces zones (voir également Harrison et Ross, 2010). Cet auteur va même jusqu'à traiter la neuroéconomie de nouvelle « néo-phrénologie » (2008b, p. 536)²⁰. Bourgeois-Gironde (2010, p. 245), de son côté, parle de stratégie de la « même zone » en insistant sur le fait que cette démarche est seulement analogique et nullement déductive.

Le raisonnement se décompose ainsi : (1) dans l'étude en question, la zone cérébrale Z est activée lorsque le sujet se livre à la tâche T ; (2) dans d'autres études, lorsque le processus cognitif X était supposé engagé, la zone Z était activée ; (3) il s'ensuit que l'activité de la zone cérébrale Z démontre que le processus cognitif X est en œuvre lorsque le sujet se livre à la tâche T dans l'étude en question. En toute rigueur, ce raisonnement n'est valide que si, dans les études antérieures, on a établi une relation *bi-univoque* entre l'engagement du processus cognitif et l'activation de la zone neuronale. Or ce n'est pas en général un résultat de cette nature dont on dispose. Outre la question de l'échelle à laquelle on devrait raisonner (le neurone ou une aire définie fonctionnellement), il semble que l'on doive admettre qu'une même zone neuronale « remplisse » souvent plusieurs fonctions distinctes : d'où l'indétermination de la valeur fonctionnelle de l'activité neuronale (Poldrack, 2006 ; Harrison, 2008b ; Phelps, 2009 ; Bourgeois-Gironde, 2010 ; Foster, 2014).

Quelles parades adopter ?

Face à cette faiblesse de la logique qui préside aux conclusions tirées des travaux d'imagerie cérébrale en neuroéconomie, la question se pose des parades à adopter. Poldrack (2006) identifie plusieurs facteurs susceptibles de conférer un certain crédit à ce raisonnement à rebours de l'activation cérébrale vers une fonction cognitive ou un état mental particulier.

Le premier facteur est la « sélectivité » de la réponse d'une région cérébrale particulière. La sélectivité d'une activation neuronale est inversement corrélée à son implication dans l'ensemble des tâches expérimentales envisageables, et donc des processus cognitifs possibles. Plus la réponse neuronale est impliquée dans un large assortiment de tâches et de processus cognitifs (*i.e.* moins elle est sélective), moins il y a de chance que l'information fournie par l'inférence inverse soit utile. L'évaluation de la sélectivité de la réponse neuronale repose en grande partie sur l'accessibilité à des banques de données (*databases*) recensant ces corrélations entre structures anatomiques du cerveau et fonctions cognitives (ce que dans le langage informatique on appelle quelquefois des « ontologies cognitives ») (Price et Friston,

²⁰ L'expression de « nouvelle phrénologie » employée par Harrison avait déjà été avancée par le psychologue William Uttal pour qualifier la démarche qui, dans les neurosciences en général, consiste à faire du niveau de l'activité cérébrale un déterminant des états mentaux à la lumière de l'imagerie cérébrale fonctionnelle (voir notamment Uttal, 2001). Foster (2014, chap.1) décortique la chaîne d'inférences sur laquelle s'appuient implicitement les expériences, des indications fournies par le débit sanguin cérébral aux détails de l'activité mentale en passant par la distribution de l'activité neuronale. Selon ce philosophe, la question est moins alors celle du bien-fondé de l'expression de « nouvelle phrénologie » que celle de la confiance que l'on peut accorder légitimement aux conclusions auxquelles parviennent ces travaux. Il énonce les conditions qui lui semblent nécessaires pour une interprétation adéquate des images du cerveau, reconnaissant que, difficiles à satisfaire, elles ne le sont pas toujours dans la littérature existante.

2005). Or les ontologies cognitives existantes ne renfermant pas d'entrées spécifiquement adaptées aux constructions théoriques économiques, les neuroéconomistes se contentent de faire référence à quelques études antérieures ayant dévoilé certaines corrélations. Et même en psychologie cognitive, discipline visée initialement par l'argumentation développée par Poldrack (2006), ces « ontologies cognitives » sont encore relativement pauvres au regard des théories psychologiques courantes – en tout cas par comparaison avec celles que l'on trouve par exemple en informatique génétique.

Un second facteur, selon l'auteur, peut intervenir dans l'évaluation de la légitimité de l'inférence inverse : celle-ci doit être d'autant plus élevée que l'activation concerne un ensemble de zones connectées en réseau. Quoique, dans le même temps, plus la région concernée est de faible taille, plus la confiance dans la réalité de l'inférence inverse doit être forte. Quelques exemples. On a déjà rappelé que l'*amygdale* (une petite zone dans le lobe temporal médian du cerveau) et l'*insula antérieure* (une zone plutôt volumineuse située dans le sillon latéral) représentaient deux zones neuronales fortement impliquées dans les comportements émotionnels. Mais elles sont particulièrement non sélectives (même l'*amygdale*, en dépit de sa petite taille), car elles remplissent de multiples fonctions. Difficile d'inférer de leur seule activation une quelconque conclusion en matière de comportement ou d'expressions émotionnelles : on ne peut observer que des corrélations (Craig, 2009). Si par exemple l'activation de l'*amygdale* est supposée témoigner de la présence d'un sentiment de peur chez le sujet (c'est l'une des émotions qui lui est souvent accolée), il sera nécessaire de posséder d'autres mesures qui confortent cette hypothèse (via par exemple une réponse électrodermale, une expression faciale ou des rapports verbaux). En revanche, pour une autre zone de faible taille comme le *noyau caudé*, une zone du striatum dont les fonctions sont plus restreintes, le raisonnement d'inférence inverse semble plus fiable²¹.

Poldrack, enfin, insiste sur le fait que les enseignements tirés des données neuronales seront d'autant plus utiles qu'ils seront confirmés par des données additionnelles d'origine différente (données comportementales ou physiologiques) (voir également Phelps, 2009).

L'une des réflexions à ce jour la plus aboutie sur la question des « inférences inverses » a été menée par Bourgeois-Gironde (2010). Ce philosophe stigmatise d'abord la démarche en fournissant plusieurs illustrations comme l'expérience de Huettel *et al.* (2006) portant sur les choix risqués/ambigus, celles de Camille *et al.* (2004) et Coricelli *et al.* (2005) consacrée au regret, ou encore celle de McClure *et al.* (2004) sur les choix intertemporels (voir notamment Serra, 2016, pour une présentation de ces expériences). Mais il suggère ensuite lui aussi quelques pistes susceptibles de donner un sens aux inférences inverses. Pour ce faire, il mobilise en particulier le concept d'« abduction » emprunté à Peirce (1878) : il s'agit d'un raisonnement visant en substance à « conclure A de l'observation de B, sachant que A implique B ». Pour autant, s'il peut se révéler utile pour générer des hypothèses, ce raisonnement ne possède pas la capacité de démontrer des propositions certaines Bourgeois-Gironde, 2010, p. 235)²². D'un autre côté, si le recours à des études antérieures peut certes aider à formuler des hypothèses interprétatives, celles-ci doivent être ensuite confirmées. En particulier la description de la fonction cognitive en jeu doit être suffisamment

²¹ Sa partie inférieure, en interaction avec le cortex ventromédian, est impliquée dans les émotions et la motivation tandis que sa partie supérieure, en interaction avec le cortex dorsolatéral, est impliquée dans le mouvement et la mémoire.

²² Dans son travail sur la logique de la science, Peirce imagine le terme d'« abduction » pour traduire un type d'inférence non déductive différente de la notion familière d'induction. L'abduction s'articule entre la déduction et l'induction de la manière suivante. La déduction et l'induction interviennent au niveau du dernier stade de l'évaluation d'une théorie : la déduction sert à tirer des conséquences testables des hypothèses explicatives que l'abduction a aidé à concevoir, et l'induction aide au final à émettre un verdict sur les hypothèses, la nature du verdict étant fonction du nombre de conséquences testables qui ont été vérifiées. Selon Peirce, l'abduction appartient à la logique parce qu'on peut en donner la caractérisation suivante : « B est observé ; mais si A était vrai, B serait une conséquence de A ; donc, il y a des raisons de suspecter que A est vrai » (voir notamment Fann, 1970).

générale pour inclure l'ensemble des autres fonctions susceptibles d'être assurées par la zone neuronale en question sans perdre sa pertinence interprétative.

Prenant acte de la difficulté qu'il y a, en neuroéconomie, à se référer à des données antérieures pour évaluer la sélectivité des réponses neuronales, Bourgeois-Gironde suggère une autre voie : il suffirait que les paradigmes expérimentaux employés dans la discipline génèrent leur propre mesure de sélectivité neuronale. Cette méthode, dite de « réduction par répétition » (réduction de la réponse neuronale observée lorsque les stimuli sont présentés plusieurs fois), a déjà été utilisée en psychologie cognitive dans le cas de la perception notamment. Mais elle n'est pas réservée à l'investigation des fonctions neuronales les plus fondamentales et pourrait être appliquée pour évaluer l'implication des processus cognitifs et affectifs pertinents en neuroéconomie.

Une réponse de nature très différente au problème de l'inférence inverse est avancée par Forest (2014, chap. 1). Elle consiste à se replier sur une conception différente de la connaissance scientifique, une épistémologie alternative qui conférerait un rôle éminent à jouer aux images cérébrales. Si l'imagerie fonctionnelle ne peut pas être source de connaissance d'un point de vue « fiabiliste » (*i.e.* si elle ne peut pas jouer le rôle d'un « indicateur fiable », un indicateur qui « dit vrai » quelles que soient les circonstances), cela ne signifie pas qu'elle ne peut pas être une source de connaissance *du tout*. Le philosophe se réfère à une conception alternative de la connaissance, l'approche « cohérentiste » : « une croyance est justifiée si et seulement si elle est en cohérence avec l'ensemble de nos croyances d'arrière-plan déjà acceptées, ou si cet ensemble peut être ajusté pour être mis en cohérence avec elle » (p. 56). La cohérence peut s'entendre d'une manière « faible », comme simple possibilité conjointe ou absence de contradiction logique, et d'une manière « forte », comme une « congruence », caractérisée par la formule suivante : « *La croyance de S selon laquelle P est une croyance justifiée si et seulement si P fait partie d'un ensemble de propositions qui se soutiennent et se confirment mutuellement* » (Lewis, 1946). Mais la congruence pose au moins deux problèmes : d'une part la sélection mutuelle des croyances ne garantit pas *ipso facto* leur validité, d'autre part rien n'est dit dans la formule en question du choix de P : pourquoi donner la priorité à P par rapport à d'autres propositions alternatives tout autant acceptables, notamment si elles ne nuisent pas directement à la cohérence des croyances acceptées ? Forest (2014, chap. 1) retient pour cette raison une version contrainte de la cohérence forte qu'il qualifie de « congruence qualifiée » : « *S est justifié lorsqu'il admet que P si et seulement s'il y a des raisons de privilégier P (plutôt que P', P''), à partir d'un ensemble de propositions qui se soutiennent et se confirment mutuellement sur la base d'enquêtes indépendantes et instructives* ». Les enquêtes doivent être « indépendantes », *i.e.* elles doivent reposer sur des méthodologies ou des sources de connaissance distinctes, et « instructives », *i.e.* elles doivent avoir fait leurs preuves en donnant accès à des classes de faits reconnues. « La congruence qualifiée permet un gain de connaissances... quand bien même nous ne disposons pas de mécanismes fiables d'acquisition d'information » (p. 58). L'interprétation d'une image peut être acceptée comme valide en se référant en particulier à sa cohérence avec les interprétations des images déjà obtenues et avec la connaissance d'arrière-plan. Selon cette perspective épistémologique, par conséquent, les références faites par les chercheurs à la littérature existante dans leur justification de l'inférence peuvent faire parfaitement sens²³.

²³ Sur la relation entre les différentes théories de la connaissance que sont le « fondationnalisme », le « fiabilisme » et le « cohérentisme », voir notamment Goldman (1988). Je rappellerai par ailleurs que de grands philosophes contemporains, comme John Rawls (1987) ou Derek Parfit (1984), ont eu également recours à l'approche cohérentiste mais dans le cadre de l'épistémologie *morale*, *i.e.* appliquée aux croyances et intuitions morales. Voir notamment Sinnott-Armstrong (1996), pour la distinction entre l'« intuitionnisme », le « constructivisme » et la « méthode de l'équilibre réflexif » en épistémologie morale en même temps que pour une discussion des liens que ces différentes méthodes de justifications des croyances morales entretiennent avec le « fondationnalisme » et le « cohérentisme » en épistémologie scientifique. On trouve

Comme on le voit, cette question des inférences inverses soulève de délicats problèmes de logique scientifique et mérite d'être traitée avec une grande attention dans les expériences qui font appel à l'imagerie cérébrale. Ce qui n'est pas forcément toujours le cas.

Quelle conclusion tirée de l'existence de ces limites inhérentes aux diverses méthodes employées en neuroéconomie ? Elles incitent sans doute à une certaine prudence quant à la portée susceptible d'être accordée aux résultats de certaines des expériences menées jusqu'à ce jour. Elles mettent pour le moins en exergue l'intérêt – voire la nécessité – de l'intégration des diverses méthodes disponibles : outils variés d'imagerie cérébrale, techniques de stimulation, manipulations pharmacologiques, études de lésions, mesures psychophysiologiques, mais aussi études neurophysiologiques avec des primates non humains.

Ernst Fehr, sans doute l'un des plus influents chercheurs européens actuellement au sein de la communauté des neuroéconomistes, affiche en particulier une telle position (Fehr, 2009). Kable (2011) et Ruff et Huetzel (2014), dans leur présentation récente de la panoplie d'outils et de méthodes dont disposent maintenant les neuroéconomistes, défendent un point de vue identique. Certaines expériences commencent d'ailleurs à associer les techniques d'intervention aux techniques d'imagerie cérébrale. Parmi elles, à titre d'exemple, l'étude par IRMf menée par Baumgartner *et al.* (2011) (à laquelle Fehr est associé) ; dans cette expérience destinée à étudier les interactions entre zones neuronales quand il existe un conflit entre une motivation égoïste et une motivation éthique, les cerveaux de certains sujets étaient scannés après déconnexion provisoire par SMT de la zone censée intervenir pour faciliter la propension à supporter le coût de décisions pro-sociales. De leurs côtés, Wong, Xue et Bechara (2011) développent plus particulièrement un plaidoyer en faveur d'une intégration systématique des mesures psychophysiologiques aux études de neuro-imagerie dans l'étude de la prise de décision en raison de l'influence déterminante qu'y tiennent les émotions. Certains neuroéconomistes, comme Paul Zak pratiquent du reste déjà un couplage simultané de mesures hormonales dans le sang avec des mesures d'activation cérébrale dans leurs expériences (par exemple, Zak *et al.*, 2005a). O'Doherty, Hampton et Kim (2007) défendent clairement quant à eux la nécessité de poursuivre en parallèle les travaux d'électrophysiologie sur les autres mammifères en raison des avantages indéniables que possède cet outil d'investigation. Plus généralement, Santos et Platt (2014) développent un plaidoyer en faveur de l'anthropologie évolutionniste, compte tenu des enseignements que cette approche est susceptible d'apporter à la connaissance de la prise de décision chez l'homme. Je relèverai enfin que participe d'un souci voisin, en imagerie cérébrale, le développement des méthodes de méta-analyses qui offrent divers moyens rigoureux pour combiner les résultats de nombreuses expériences différentes (Wager, Lindquist et Kaplan, 2007 ; Eickhoff *et al.*, 2012 ; Yarkoni *et al.*, 2011).

En définitive, ce n'est qu'en croisant les diverses techniques puisées dans la riche et diversifiée boîte à outils désormais accessible et en comparant les différentes méthodes d'investigation envisageables que l'on améliorera la crédibilité des résultats. À cet égard, on peut regretter avec Kable (2011) une trop forte proportion d'études dont les résultats s'appuient encore aujourd'hui sur la seule interprétation de données d'imagerie cérébrale dans les travaux de neuroéconomie – même s'il est vrai qu' on assiste ces dernières années à une multiplication des études recourant à la neuroendocrinologie.

dans Clément, Le Clainche et Serra (2008) des précisions relatives au rattachement des grandes théories économiques de la justice modernes aux diverses conceptions de l'épistémologie morale.

2. LA LÉGITIMITÉ DE LA DÉMARCHE ?

La neuroéconomie soulève également un problème d'ordre exclusivement méthodologique voire épistémologique. La critique déborde d'ailleurs le périmètre strict de la neuroéconomie et concerne aussi l'économie comportementale lorsqu'elle recueille des données cognitives et non comportementales, *i.e.* qui portent sur les états et processus cognitifs supposés responsables des comportements et non sur les comportements eux-mêmes. La très grande majorité des données employées en économie sont des données comportementales, qu'elles soient recueillies sur le terrain ou dans un environnement expérimental contrôlé. C'est à partir de ces données que l'on teste les modèles, que l'on mesure les variables subjectives, et ce sont elles que l'on prédit à partir des modèles. Mais dans certains travaux d'économie comportementale (quand on utilise les méthodes de mesure ou d'induction des émotions)²⁴ et plus particulièrement en neuroéconomie, on recueille aussi des *données physiologiques et neuronales*. Il existe aujourd'hui un désaccord concernant la question de savoir si ces données non comportementales sont « pertinentes » ou non pour l'économie. Il s'agit de s'interroger, plus précisément, sur la nature des rapports épistémologiques entre les données non comportementales et les modèles de choix utilisés en économie, qu'il s'agisse des modèles standards ou des nouveaux modèles alternatifs élaborés en économie comportementale, comme la théorie des perspectives (Kahneman et Tversky, 1979 ; Tversky et Kahneman, 1992). Une majorité des positions et des arguments avancés dans le débat constituent des réactions à la thèse de l'« économie sans esprit » (*mindless economics*) défendue par Gul et Pesendorfer (2005/2008) (voir notamment les articles rassemblés dans Caplin et Schotter, 2008, et dans les numéros spéciaux de la revue *Economics and Philosophy* (2008/24) et du *Journal of Economic Methodology* (2010/17(2))).

2.1. L'absence de pertinence des données non comportementales pour tester des modèles de choix économiques : la thèse de l'« économie sans esprit » (*mindless economics*)

Sans rejeter l'éventuelle *fécondité heuristique* des enseignements que peuvent fournir les données non comportementales, Gul et Pesendorfer (2005/2008) nient que ces données puissent avoir une *pertinence épistémologique* pour les modèles de choix de l'économie. Selon ces auteurs, les données non comportementales ne peuvent entretenir des relations « confirmationnelles » avec les modèles de choix, *i.e.* elles ne peuvent ni confirmer ni infirmer les modèles de choix, qui sont en quelque sorte immunisés de toute remise en question via ces données. Pour ces auteurs, selon la formule employée par Cozic (2012, p. 138), les modèles de choix économiques sont « cognitivement muets ». Ce mutisme cognitif des modèles de choix signifie en particulier qu'ils n'ont pas de *conséquences* pour les données cognitives, *i.e.* il est certes possible d'en effectuer une lecture mentaliste, mais ils ne contraignent nullement les données cognitives que l'on peut recueillir²⁵.

Exprimée autrement, la thèse développée par Gul et Pesendorfer peut se résumer ainsi : l'économie ne s'intéresse pas aux mêmes questions que les neurosciences ou la psychologie. Elle poursuit des objectifs différents et doit dès lors employer des abstractions propres à ses finalités. Les variables biologiques ou psychologiques qui causent les choix ne sont pas pertinentes d'un point de vue économique. L'économiste n'a pas besoin de savoir, par exemple, si le fait de préférer x à y « est motivé par la recherche du bonheur, le sens du devoir, une obligation religieuse ou une impulsion » (p. 24). Seul importe pour lui les relations de préférence *révélées par les choix*. Cette défense d'une approche strictement instrumentaliste n'implique

²⁴ Voir notamment Petit (2015).

²⁵ Cozic (2012) attribue deux significations au mutisme cognitif des modèles de choix économiques : (i) ils ne *contiennent* pas de concepts cognitifs ; (ii) ils n'ont pas de *conséquences* pour les données cognitives. Mais il prétend de manière convaincante que la première implication n'est guère recevable, sauf à accepter sans réserve la sémantique de la préférence révélée.

pas pour autant que l'économie soit entièrement isolée de la recherche psychologique. Ce que cela implique c'est qu'elle ne doit intégrer que les *données pertinentes*, *i.e.* les données sur les choix et les comportements. À cet égard, les données produites par l'économie ou la psychologie expérimentales peuvent très bien être utilisées pour calibrer un modèle économique ou prédire des choix ou des équilibres futurs. En revanche, les données ou les variables non reliées aux choix, telles les données relatives aux activités neuronales, n'ont pas à être intégrées dans le modèle. Quel crédit accorder à cette critique de Gul et Pesendorfer à l'encontre de la neuroéconomie ?

2.2. La portée de la critique de Gul et Pesendorfer

Plusieurs réponses à la critique de Gul et Pesendorfer (2005/2008) ont été avancées. Examinons quelques-unes de ces réponses en les ordonnant selon la distance croissante qu'elle manifeste vis-à-vis de la méthodologie dominante en économie.

Une première attitude revient à nier le problème en minimisant les différences qui existeraient entre la théorie économique standard, l'économie comportementale et la neuroéconomie. Camerer (2008b), notamment, présente ces deux dernières comme de simples *généralisations* de la théorie du choix rationnel : alors que l'économie comportementale ajoute des données et des variables *psychologiques* à la théorie, la neuroéconomie élargit encore l'analyse en dotant le comportement de composantes *biologiques*. Elle rajoute une composante mécaniste à la théorie économique standard sans que l'approche ne soit fondamentalement différente. Selon cette perspective, l'économie comportementale et la neuroéconomie permettraient de généraliser « la théorie du choix rationnel de façon à y insérer formellement des limites à la rationalité, à la volonté et à l'intérêt personnel » (Camerer, 2008b, p. 44). La question centrale (à laquelle Gul et Pesendorfer ne répondent pas) est de savoir si cette généralisation permet « d'améliorer la capacité à comprendre et prédire les choix, tout en maintenant la discipline mathématique et le recours aux données comportementales » (p. 45).

Wilkinson (2008), quant à lui, évoque une certaine incompréhension courante des ambitions de l'économie comportementale et de la neuroéconomie. L'instrumentalisme (en substance, la méthode du « comme si »), largement dominant en économie, est tout à fait acceptable – même par un auteur comportementaliste –, à *condition* toutefois que la théorie fournisse des prédictions *correctes*. En neuroéconomie, on ne prétendrait pas ordinairement que les données neuronales puissent *falsifier* une théorie. L'argumentation est la suivante : puisque les théories économiques standards sont souvent falsifiées par des études empiriques s'appuyant sur des données reliées à la sémantique de la préférence révélée (*i.e.* elles sont falsifiées en leurs propres termes), les travaux de neuroéconomie peuvent se révéler utiles pour comprendre *pourquoi* les théories sont invalidées, en recourant aux processus psychologiques et biologiques sous-jacents. En conséquence de quoi, ces études peuvent servir à établir que certaines hypothèses implicites du raisonnement « comme si » du modèle standard sont de mauvaises hypothèses dans la mesure où elles conduisent à des prédictions fausses, suggérant du coup des pistes susceptibles de conduire à l'élaboration de « meilleurs » modèles économiques. Ainsi, grâce à une bonne connaissance des mécanismes qui régissent les décisions, les prédictions théoriques peuvent être améliorées. Si la neuroéconomie ne permet pas de tester *directement* les modèles économiques courants, elle doit malgré tout être considérée comme une approche pertinente pour évaluer et tester indirectement la plausibilité des conjectures économiques.

On peut du reste prétendre avec Aydinonat (2010) que la pertinence de la neuroéconomie est sans doute encore plus large : au-delà du rôle indirect qu'elle est susceptible de jouer dans la validation empirique des modèles économiques courants, elle devrait être jugée aussi sur sa capacité à améliorer les *explications* de

certaines phénomènes économiques particuliers, pour lesquels il s'avère souvent nécessaire de prendre en compte certaines de leurs caractéristiques spécifiques qui imposent d'envisager des facteurs supposés en dehors du cadre théorique courant de l'économie. Par exemple, le « contexte » d'une décision devrait être pris en compte, parce que la manière dont le cerveau évalue les récompenses dépend de l'histoire personnelle de l'individu et du contexte en jeu. Plus largement, la neuroéconomie devrait contribuer à améliorer notre *compréhension*²⁶ des phénomènes économiques en général. Aydinonat prend l'exemple de l'illusion monétaire, dont l'existence peut s'expliquer par le fait que l'argent ne serait pas perçu par le cerveau comme un simple instrument qui rend des services mais aussi comme une sorte de « drogue » qui « s'empare » du système de récompense de diverses façons. Plusieurs expériences de neuroéconomie semblent suggérer en effet que l'impact de l'argent dans le cerveau est comparable à celui d'une drogue. Pour le cerveau, l'argent ne serait pas qu'un moyen permettant d'accéder à certaines consommations, il serait source de plaisir en soi. C'est pourquoi la « consommation » de l'argent s'avère souvent très addictive, à l'image des « renforcements » primaires comme la drogue, le sexe ou la nourriture (Delgado *et al.*, 2000 ; Knutson *et al.*, 2001). (Bourgeois-) Gironde (2008, chap. 2) discute plusieurs expériences traitant de cette question du pouvoir addictif de l'argent en même temps que de la motivation à l'obtenir et de la frustration qu'il génère en cas de manque.

De son côté, Harrison (2008a), bien que plutôt critique à l'égard des expériences de neuroéconomie – en tout cas, telles qu'elles étaient conçues et mises en œuvre à l'origine (voir *supra* section 1) –, n'adhère pas pour autant à la thèse de Gul et Pesendorfer. Si sa position est pour le moins mitigée quant à la fiabilité des résultats en relation avec les limites attachées à l'expérimentation en laboratoire et les difficultés d'interprétation des données neuronales, il rejette l'argument du manque de légitimité scientifique. Le rôle potentiel des données neuronales ne doit pas être écarté *a priori*. Il conviendrait cependant de reformuler au préalable les mécanismes formels qui régissent la prise de décision économique en plaçant au premier plan les *processus* qui conduisent au choix plutôt que le choix lui-même. Selon Harrison, la neuroéconomie *pourrait* prétendre « apporter des réponses à de nouvelles questions » (faisant allusion au propos de Smith, 2008), à condition de « concevoir formellement le comportement économique comme le résultat de processus algorithmiques... L'idée est déjà familière à de nombreux économistes comportementalistes, mais uniquement comme métaphore » (p. 330). Les données neuronales, au sein de ce projet, auraient à coup sûr un rôle à jouer, mais pour l'heure la priorité serait de s'engager dans un véritable travail de communication entre l'économie et la psychologie cognitive. C'est la voie dans laquelle s'oriente justement la recherche ces dernières années, une évolution qui est amorcée parallèlement en économie expérimentale par Vernon Smith, l'un des pionniers de cette discipline (Smith, 2003, 2008). Après avoir marqué à l'origine les spécificités de la méthode expérimentale en économie au regard des expériences menées antérieurement par les psychologues, cet économiste a engagé par la suite un travail de réorientation de la discipline vers une connexion plus profonde avec la psychologie (Guala, 2008 ; Serra, 2012a, 2015, et Cot et Ferey, 2016, proposent des ébauches de l'histoire de l'économie expérimentale).

Dans un ordre d'idées voisin, Walliser (2012) qualifie de « position intégriste » le point de vue défendu par Gul et Pesendorfer en ce sens qu'il reviendrait à postuler que l'étude des choix au seul niveau comportemental ne serait nullement restrictive, que la connaissance de leurs bases mentales et neuronales serait sans intérêt. À l'inverse, une « position œcuméniste » – qui a ses faveurs – consiste à accepter de prendre en compte aussi les états mentaux et neuronaux « observés ».

²⁶ Une *explication* est une réponse à une question très spécifique relative à un ensemble circonscrit d'aspects du phénomène en jeu plutôt qu'à d'autres aspects. La *compréhension* d'un phénomène nécessite de posséder des informations portant sur un ensemble beaucoup plus large de ses aspects et sur d'autres relations de dépendance pertinentes, elle permet de répondre à des questions qui vont au-delà du questionnement : « si...alors ».

Une solution plus radicale est soutenue par Quartz (2008). Traçant un parallèle avec les controverses qui eurent lieu au sein de la science cognitive vers le milieu des années 1980 lors de l'émergence des neurosciences cognitives et computationnelles, il souligne aussi le danger attaché à la thèse d'une autonomie stricte de l'économie vis-à-vis des neurosciences. On ne voit pas en effet, selon lui, ce qui pourrait justifier l'argument de Gul et Pesendorfer selon lequel les neurosciences ne pourraient réfuter les modèles économiques au prétexte qu'ils ne posent aucune hypothèse et ne tirent aucune prédiction de nature physiologique – si ce n'est la *thèse de l'autonomie*. Or, selon l'auteur, si la théorie économique ne pose aucune hypothèse sur la physiologie du cerveau, ce n'est pas vrai qu'elle ne puisse effectuer des prédictions susceptibles d'être confirmées ou réfutées par les neurosciences. Se référer à cet égard aux interprétations que donnent certains auteurs des résultats de leurs expériences : les données « confirmeraient » le modèle espérance-variance plutôt que le modèle de l'utilité espérée (Quartz, 2008, et Schultz *et al.*, 2008, à propos notamment de l'expérience de Preuschoff, Bossaerts et Quartz, 2006), les données « parleraient en faveur » d'un modèle de choix incertain qui évalue séparément l'utilité espérée et l'ambiguïté (Camerer, 2008b, à propos de l'expérience de Hsu *et al.*, 2005), elles « contrediraient » la conséquence des modèles de choix en incertitude qui suppose que les processus en œuvre dans les choix risqués sont subordonnés à ceux qui interviennent dans les situations incertaines (Huettel *et al.*, 2006), elles « confirmeraient » le modèle quasi hyperbolique et seraient « en défaveur » du modèle de l'escompte exponentiel (McClure *et al.*, 2004a), ou encore les résultats seraient considérés comme une « preuve à l'appui » du modèle des préférences sociales (de Quervain *et al.*, 2004 ; Sanfey *et al.*, 2003) (toutes ces expériences sont évoquées succinctement dans Serra, 2016, chap. IV). *In fine*, on serait donc ramené à la querelle épistémologique récurrente entre les tenants d'une position « réaliste » (en bref, le rôle d'une théorie est de postuler des termes susceptibles de correspondre à des entités du monde réel) et les défenseurs d'une position « instrumentale » (en bref, seule importe la capacité prédictive de la théorie)²⁷. Invoquant les désillusions auxquelles auraient conduit les interprétations instrumentalistes du niveau symbolique de la science cognitive (Dennett, 1987 ; Baker, 1989) et, plus généralement, l'entrave au progrès instaurée par cette thèse de l'autonomie de la science cognitive, Quartz rejette la thèse de l'« économie sans esprit » défendue par Gul et Pesendorfer, en raison de l'« isolationnisme » et du manque de flexibilité méthodologique qu'elle incarne.

Reste un dernier argument d'ordre fort différent à l'encontre de la thèse avancée par Gul et Pesendorfer. Il est formulé par Caplin et Dean (2009, p. 31) qui observent que l'approche *axiomatique* originale qu'ils préconisent se trouve *a priori* immunisée contre la critique formulée par ces auteurs puisqu'elle donne naissance à des modèles compatibles formellement avec n'importe quel type de données, *quelle que soit leur origine* (voir Caplin et Glimcher, 2014, pour une présentation de la « neuroéconomie axiomatique »)²⁸

Face à ces multiples arguments avancés en réponse à la critique émise par Gul et Pesendorfer (2005/2008) à l'endroit de la neuroéconomie, doit-on conclure que la difficulté soulevée par ces auteurs peut être aisément contournée ? Pas forcément. Car, d'un strict point de vue épistémologique, le problème

²⁷ Sur la méthodologie de l'économie positive de Friedman et sa relation avec l'instrumentalisme, voir Mäki (2003). On peut trouver une discussion du réalisme, de l'instrumentalisme et d'autres formes d'antiréalisme en philosophie des sciences dans Bird (1998, chap. 4) et Chalmers (1999, chap. 15). L'adoption d'une position réaliste de la science ne constitue pas la seule base envisageable d'une argumentation critique à l'égard de la méthodologie prônée par Friedman. Hausman (1992, chap. 9) développe une critique alternative qui consiste à nier que l'argument de la « non pertinence du réalisme des hypothèses » serait impliqué par la conception instrumentaliste selon laquelle les théories devraient être évaluées par le succès de leurs prédictions.

²⁸ Comme les économistes qui, depuis Samuelson et Houthakker, prétendent que dire qu'un individu maximise une fonction d'utilité est la même chose que dire qu'il obéit aux fameux axiomes de la théorie de la préférence révélée, ces auteurs se demandent si l'on ne pourrait pas identifier un ensemble restreint d'axiomes qui caractériserait toutes les classes d'algorithmes d'« erreur de prévision de la récompense » (*reward prediction error*), i.e. le mécanisme censé à l'œuvre dans le système d'apprentissage de la récompense qui caractérise le nouveau paradigme adopté par les neurosciences dans les années 1990 (voir plus loin).

reste entier. Les modèles usuels de choix en économie, c'est un fait, ne se prononcent pas sur les mécanismes physiologiques ou neurologiques en jeu lors de la prise de décision. Il manque des *hypothèses de liaison* (au sens de Teller et Pugh, 1983) entre les modèles économiques de choix et les données cognitives ou neuronales ; on se trouve confronté à un problème de « liaisons manquantes » (Cozic, 2012, p. 141). L'argument fort qui plaide en faveur de la thèse de Gul et Pesendorfer s'appuie de fait sur la constatation que les seules variables explicitement prises en compte dans les modèles courants de choix économiques sont bien de nature comportementale. Ils ne mentionnent guère de variables observables en dehors des comportements. Certes, cela n'exclut pas que l'on puisse inférer des modèles certaines conséquences en posant quelques hypothèses générales plausibles relatives à la cognition. Pour autant, « il semble difficile de trouver des conséquences de modèles de choix en matière de données cognitives qui soient à la fois générales et précises » (Cozic, 2012, p. 145). C'est sans doute l'une des raisons principales qui a nourri le doute exprimé quelques années en arrière par certains économistes à l'égard de la pertinence épistémologique de la neuroéconomie (Bernheim, 2009 ; Harrison, 2008a ; Rubinstein, 2008, notamment²⁹).

De ce point de vue, quelques pistes ont été explorées pour suggérer l'existence d'hypothèses de liaison entre les modèles de choix économiques et les données cognitives et tenter d'attribuer une certaine pertinence « confirmationnelle » aux données cognitives. Ainsi Quartz (2008), qui invoque une distinction entre les conséquences cognitives « implicites » et « explicites » d'un modèle de choix³⁰, critiqué en cela par Cozic (2012), qui préfère voir dans certaines « hypothèses auxiliaires plausibles », rajoutées aux modèles de choix, la liaison qu'ils entretiennent avec les données cognitives. Mais dans ce cas, comme le reconnaît ce philosophe, les relations confirmationnelles deviennent contingentes : elles dépendent de facteurs extrinsèques au modèle (*i.e.* les hypothèses auxiliaires).

Explorant la situation type d'une expérience de neuroéconomie par imagerie cérébrale dans le cas d'un problème de choix binaire en situation certaine, Cozic montre qu'en introduisant une hypothèse auxiliaire H de nature psycho-cérébrale, selon laquelle l'activité cérébrale dans l'aire neuronale A encode les préférences de l'individu (hypothèse raisonnable, car elle découle d'une autre hypothèse constitutive du modèle élémentaire de choix : l'individu choisit ce qu'il préfère), on dispose de conséquences observables portant sur des données neuronales susceptibles de *confirmer* le modèle élémentaire de choix (en même temps que l'hypothèse auxiliaire) : on observe la trace neuronale du fait que l'individu choisit l'option préférée. En revanche, l'*infirmité* du modèle ne va pas de soi : doit-on rejeter l'hypothèse selon laquelle l'individu choisit l'option qu'il préfère ou l'hypothèse auxiliaire H, selon laquelle l'activation de l'aire A encode les préférences ? La réponse à cette question n'est pas triviale, sauf pour les tenants de la théorie de la préférence révélée, qui rejettent toujours l'hypothèse H plutôt que l'hypothèse de choix préférentiel. Dans le cas contraire, en revanche, on peut très bien envisager qu'il existe des raisons de préserver l'hypothèse H et d'admettre qu'il est probable que l'on ait infirmé l'hypothèse testée du choix préférentiel. Or, on sait qu'il existe en épistémologie de fortes objections à l'égard de la théorie de la préférence révélée (voir notamment Hausman, 2000, 2008).

La ligne argumentative développée ici revient somme toute à prétendre que la principale difficulté soulevée par Gul et Pesendorfer (2005/2008) aurait tendance à se dissiper à partir du moment où l'on ne pose pas

²⁹ Cette difficulté pour établir des « hypothèses de liaison » entre les différents niveaux en jeu dans l'analyse n'est pas l'apanage de la neuroéconomie. Elle est présente dans toutes les sciences cognitives. Ainsi, la question épistémologique de l'articulation entre le niveau cérébral et le niveau mental se pose-t-elle toujours entre autres en psychologie cognitive (voir notamment Henson, 2005).

³⁰ Il développe cette argumentation pour défendre le point de vue selon lequel les résultats de certaines expériences sur les choix financiers « confirmeraient » le modèle espérance-variance parce qu'ils « vérifient » certaines de leurs conséquences cognitives « implicites » (voir Serra, 2016, chap. IV).

comme dogme la doctrine de la préférence révélée. La thèse de l'« économie sans esprit » ne serait au fond qu'un plaidoyer pour défendre ce fonds commun méthodologique de l'économie depuis le milieu du 20^e siècle³¹.

3. L'UTILITÉ DES RÉSULTATS POUR L'ÉCONOMISTE ?

La neuroéconomie permet de révéler les composantes et les processus physiologiques et neuronaux impliqués dans la prise de décision économique. À ce titre, toutefois, le problème serait que les phénomènes dont la neuroéconomie révèle les fondements neuronaux sont généralement déjà bien identifiés sur le plan comportemental. C'est là un autre argument qui a nourri un temps le « neuroscepticisme » de certains économistes (Smith, 2008, chap. 14 ; Harrison, 2008a, 2008b ; Rubinstein, 2008, notamment). Ainsi, pour Harrison (2008a), les phénomènes psychologiques sur lesquels insistent les neuroéconomistes « sont bien sûr connus des économistes depuis des décennies » (p. 308). Si un phénomène est déjà bien cerné aux niveaux psychologique et comportemental, qu'apporte à l'économiste la connaissance des corrélats ou des mécanismes neuronaux ?³² D'autant que le coût de certaines de ces expériences de neuroéconomie (dans le cas de la neuro-imagerie en particulier) est sans commune mesure avec celui des expériences menées en économie comportementale. Rubinstein (2008), en particulier, développe une critique de cette nature en préconisant de restreindre l'usage des outils de neuro-imagerie, dont le coût très élevé inciterait les chercheurs à surestimer leur fiabilité, au profit des indicateurs simples et peu coûteux empruntés aux psychologues (oculométrie, temps de réaction, réaction électrodermale, etc.)³³.

Que penser de cette thèse qui, somme toute, revient à prétendre que les expériences de neuroéconomie n'apporteraient que peu d'enseignements réellement bénéfiques au plan économique ou encore se révéleraient de peu d'intérêt d'un point de vue pragmatique ? Elle semble aujourd'hui excessive sinon injustifiée, en tout cas si l'on n'assimile pas la neuroéconomie à la seule « économie comportementale dans le scanner » (*behavioral economics in the scanner*) – le programme développé à l'origine par certains économistes – mais que l'on s'attache aussi à l'« économie neuronale » (*neural economics*) – le programme défendu initialement par certains neurobiologistes³⁴. Evoquons succinctement au préalable les arguments traditionnellement avancés par certains économistes en faveur de la neuroéconomie.

³¹ À cet égard, ce débat autour de la légitimité scientifique de la neuroéconomie (et plus généralement de l'économie comportementale) n'est pas sans rappeler la controverse entre Gary Becker et Edward Wilson lors de la naissance de la sociobiologie dans les années 1970 (voir notamment Wilson, 1975 et Becker, 1976).

³² Remarquons que, du point de vue des neuroscientifiques, on peut prétendre à rebours que dans une perspective réductionniste le projet de la neuroéconomie ne sera vraiment fructueux que s'il s'appuie justement sur des théories économiques et psychologiques déjà bien établies *en leurs propres termes* (i.e. par des variables de nature économique et mentale respectivement) (Gallistel, 2009).

³³ Selon Rubinstein (2008), on doit malgré tout s'attendre à ce que la neuroéconomie influence l'économie parce que l'économie est avant tout une culture et non une science, en entendant par « culture » une collection d'idées admises et de conventions utilisées par notre cerveau.

³⁴ Deux orientations majeures sont en effet identifiables au sein de la neuroéconomie. Selon une première perspective, elle est présentée comme un domaine transdisciplinaire mobilisant les outils des neurosciences pour identifier les bases neuronales des décisions économiques. Elle vise simplement à compléter la démarche qui caractérise l'économie comportementale en s'interrogeant sur les origines biologiques des présupposés relatifs aux caractéristiques psychologiques des individus dans les modèles que propose cette approche des phénomènes économiques (McCabe, 2003 ; Zak, 2004 ; Camerer, Loewenstein et Prelec, 2005 ; Sanfey et al., 2006 ; Camerer, 2007). C'est le philosophe Don Ross qui, en qualifiant d'« économie comportementale dans le scanner » ces travaux, distingue un autre courant qu'il nomme l'« économie neurocellulaire » (*neurocellular economics*) (Ross, 2005, 2008), que je préfère qualifier d'« économie neuronale » (*neural economics*) en suivant la suggestion de Montague et Berns (2002). Il s'agit cette fois d'emprunter à la microéconomie ses concepts et ses mathématiques de l'optimisation et de l'équilibre pour modéliser l'activité des cellules cérébrales et tester notamment certaines hypothèses portant sur l'apprentissage neuronal. Voir également Vromen (2007, 2011) pour l'identification de ce clivage désormais largement reconnu. Dans leur introduction à la 1^{ère} édition du *Handbook* de neuroéconomie (Glimcher et al., 2009), les éditeurs ne cachaient d'ailleurs pas cette division interne au sein de la communauté des neuroéconomistes. Cette distinction est utile – ne serait-ce que parce certaines critiques adressées à la neuroéconomie visent surtout le premier courant. Cependant, elle perd progressivement de sa pertinence à la lueur de l'évolution récente des

3.1. Pourquoi les économistes ont-ils besoin de la neuroéconomie ?

C'est un fait que l'économie comportementale, en quelque décennies, a su produire un certain nombre de modèles alternatifs aux modèles de l'économie standard, que ce soit dans le domaine des choix individuels ou dans celui des choix en situation d'interaction (voir notamment Camerer, Loewenstein, Rabin, eds., 2003, Camerer, 2003, Wilkinson, 2008, Gilboa, 2009, Heukelom, 2014, Serra, 2017). Ces succès sont dus à la fois à des avancées théoriques et au recours à l'expérimentation dans la production de données (Fudenberg, 2006 ; Pesendorfer, 2006). Pourquoi, dans ces conditions, l'économie comportementale aurait-elle besoin de ce « complément » que représente la neuroéconomie ?

Une première réponse réside dans la richesse et la puissance de ces nouveaux outils d'investigation dont disposent dorénavant les spécialistes des neurosciences. On attend en particulier de ces outils qu'ils permettent d'« affiner les interprétations psychologiques » que les modèles de l'économie comportementale fournissent des comportements observés (Camerer, Loewenstein et Prelec, 2005). Cette position, exprimée par trois promoteurs enthousiastes de l'approche neuroéconomique, semble partagée par la plupart des économistes qui y participent mais aussi par certains économistes de premier plan moins engagés dans cette nouvelle démarche scientifique, tel le prix Nobel d'économie, Jean Tirole, qui voit dans la neuroéconomie un instrument destiné à « affiner les modèles » (Tirole, 2007).

Grâce à l'utilisation de ces nouveaux outils, la neuroéconomie pourrait permettre aussi de s'assurer que les modèles de l'économie comportementale « sont sur la bonne voie », en vérifiant par exemple si des variables distinctes correspondent à des composantes ou à des processus distincts. Elle pourrait apporter des confirmations aux explications que fournit l'économie comportementale aux écarts entre les résultats expérimentaux et la théorie économique standard (Camerer, 2008b).

Mais une autre réponse peut être avancée : elle renvoie à l'existence de ces effets pervers portant sur la relation entre l'observation et le dispositif d'observation auxquels sont confrontées les expériences économiques en laboratoire, comme le « biais de l'expérimentaliste » (appelée encore « effet d'attente », « effet de demande » ou effet Hawthorne), l'« effet de réputation » ou l'« effet du volontariat ». Voir notamment Serra (2012b) pour une discussion de ces différents biais et des méthodes employées par les économistes pour parer à certains d'entre eux, comme la pratique de la rémunération des sujets en liquide en fin d'expérience et l'usage du protocole en « double anonymat » ou encore l'usage de la tromperie dans les expériences menées par les psychologues (pratique qui est en revanche à proscrire pour les économistes). C'est là un autre raison émise souvent par certains économistes pour justifier le recours dans leurs expériences à l'emploi des outils développés par les neurosciences. En enquêtant directement sur le cerveau, sur les activités cérébrales qui sont à l'origine des comportements, les outils des neurosciences permettent des analyses beaucoup plus objectives, non polluées par ces effets indésirables qui peuvent être en jeu quand on enquête sur la personne (Camerer, Loewenstein et Prelec, 2004 ; Kenning et Plassman, 2005)³⁵.

recherches. Car on peut prétendre que ces deux grandes orientations historiques de la neuroéconomie sont en passe de converger vers l'adoption d'un agenda de recherche commun à une large fraction de la communauté (voir Serra, 2016).

³⁵ Pour certains auteurs, cette vertu de la neuroéconomie resterait sujette à caution dans les expériences qui utilisent certains outils comme l'oculométrie (comment s'assurer que le sujet ne va pas contrôler consciemment son regard ?) ou encore les réponses électrodermales (les sujets auraient la possibilité de contrôler consciemment leur rythme cardiaque – et donc indirectement la sudation) (Pelloux, Rullière et Van Winden, 2009, p. 16). Ce risque me paraît toutefois assez faible. Pour un plaidoyer en faveur de l'emploi de l'oculométrie, des réponses électrodermales et du contrôle du rythme cardiaque dans les expériences de neuroéconomie, voir respectivement Glanholt et Reingold (2011), Dawson, Schell et Courtney (2011) et Dulleck *et al.* (2011).

Un dernier argument fort en faveur de la neuroéconomie ne peut être ignoré ; il renvoie au fondement même du choix rationnel en économie. Grâce à la neuroéconomie, l'affirmation selon laquelle le choix résulte de la maximisation de la valeur que l'agent attribue aux différentes options en lice dans un problème de décision devient potentiellement falsifiable et donc vraiment scientifique. Comme chez les économistes, la notion de valeur subjective d'une décision est un concept central en neuroéconomie. Mais il existe une différence cruciale. En économie, le concept de valeur est comportemental et analytique, non psychologique. L'argument du « comme si » souffre d'une limite fondamentale : sur la base du seul comportement, les valeurs ne peuvent être mesurées indépendamment du choix. L'affirmation selon laquelle « le choix maximise la valeur » est circulaire. La neuroéconomie brise cette circularité en prouvant que les valeurs sont bien calculées par le cerveau. En établissant la correspondance entre un signal neuronal et une mesure comportementale de la valeur d'une décision, ce signal observable dans le cerveau grâce à l'imagerie cérébrale fournit une mesure indépendante de la valeur, en principe dissociable des choix (Padoa-Schioppa, 2011 ; Padoa-Schioppa et Conen, 2017).

3.2. Une appréciation d'ensemble de la critique

Que l'on soit convaincu ou non par ces arguments avancés en faveur de la neuroéconomie, comment recevoir la critique selon laquelle les expériences neuroéconomiques n'apporteraient au fond rien d'essentiel à la connaissance des comportements économiques ? Si l'on veut répondre à cette question, il faut s'intéresser aux résultats de ces expériences et aux enseignements que l'on peut en tirer. Il n'est pas question d'évoquer ici dans le détail les premiers enseignements de la neuroéconomie (voir notamment Glimcher *et al.*, eds., 2009 ; Glimcher et Fehr, eds., 2014 ; Serra, 2016, chap. 4). Je me contente d'évaluer la critique en me référant aux plus significatifs d'entre eux, en distinguant les deux programmes historiques, l'économie comportementale dans le scanner et l'économie neuronale, en même temps que les évolutions récentes.

L'économie comportementale dans le scanner

On doit d'emblée reconnaître que, s'agissant de certaines des expériences de première génération du programme de l'économie comportementale dans le scanner ce jugement est en partie fondé. Car les résultats de ces expériences n'ont effectivement que corroboré des explications en termes d'états ou processus mentaux déjà avancées dans certains contextes par l'économie comportementale. Comme le regrettait Vernon Smith, quelques années en arrière, « initialement, les nouveaux outils ont tendance naturellement à être appliqués à d'anciennes questions » (Smith, 2008, p. 313). Ainsi, la théorie des perspectives (*prospect theory*), en tant qu'alternative principale à la théorie de l'utilité espérée proposée en économie comportementale (Kahneman et Tversky, 1979 ; Tversky et Kahneman, 1992), a-t-elle fait l'objet de multiples expériences de neuroéconomie qui ont confirmé certaines de ses hypothèses centrales comme l'aversion aux pertes, l'existence d'un point de référence dans les jugements probabilistes et la non-linéarité des jugements probabilistes (voir Fox et Poldrack 2014, ainsi que Louie et De Martino, 2014, pour la dépendance au contexte plus particulièrement)³⁶. De même, bon nombre d'explications d'ordre

³⁶ Kahneman a participé lui-même à l'une des premières expériences de neuroéconomie (Breiter *et al.*, 2001). Le statut des données neuronales dans cette expérience, toutefois, est ambigu. Comme toutes les expériences pionnières du tournant des années 2000, l'expérience prétend s'inscrire dans le cadre théorique de l'apprentissage de la récompense ; or il est clair que la théorie des perspectives y joue aussi le rôle de paradigme expérimental. Les données neuronales sont tour à tour considérées comme des paramètres du modèle de Kahneman-Tversky (des variables *exogènes* qu'il est nécessaire d'estimer pour « calibrer » le modèle) et des variables explicatives (des variables *endogènes* qui sont corrigées progressivement par un processus neuronal d'apprentissage). Cette expérience illustre clairement la difficulté à laquelle on est confronté lorsqu'on transpose dans le cadre théorique de l'apprentissage de la récompense l'interprétation comme « anomalies » des écarts entre rationalité économique « idéale » et comportements observés. En neurobiologie, c'est par référence à la dynamique de l'apprentissage que l'irrationalité dans le comportement s'évalue (Fox et Poldrack, 2014).

psychologique avancées en économie comportementale pour comprendre les soi-disant « anomalies » observées dans les situations d'interaction sociale (voir notamment Serra, 2017, chap. 3) ont été confirmées par de nombreuses expériences de neuroéconomie sociale (on peut trouver des synthèses dans Sanfey et Rilling, 2011, Fehr et Krajbich, 2014, Declerck et Boone, 2016, Serra, 2016, chap 4, Dreher et Tremblay, 2017, Part III).

Ce constat, malgré tout, mérite d'être nuancé. On peut objecter en effet que non seulement ces hypothèses acquièrent du coup une justification *objective* plus solide, mais il se trouve en outre que certaines de ces expériences commencent à apporter des éclairages inédits sur des questions d'intérêt pour les économistes dans maints domaines. En dévoilant les bases neuronales d'éléments spécifiques du choix aussi variés que le désir, le regret, l'anticipation ou l'attitude à l'égard du risque, de l'ambiguïté ou du temps, le programme de l'économie comportementale dans le scanner participe à une meilleure compréhension des motivations propres à la décision individuelle. Certains travaux, par exemple, explorent la différence entre déception et regret. Le regret engendre des réponses physiologiques plus intenses que la déception. Plus exactement les deux émotions paraissent se cumuler : on est triste du résultat atteint et triste d'avoir opté pour un mauvais choix (Camille *et al.*, 2004). D'autres expériences apportent des explications plus précises que l'économie comportementale sur l'aversion à l'égard des pertes ou encore l'aversion à l'égard de l'ambiguïté, deux émotions courantes fréquemment observées dans les décisions financières en particulier. L'aversion à l'égard des pertes est expliquée par l'interaction entre les structures neuronales impliquées dans l'anticipation, l'encodage et le calcul de l'effet hédonique d'une décision risquée. On manifeste de l'aversion aux pertes parce que l'on anticipe l'impact émotionnel négatif d'une perte. La neuroéconomie jette également un éclairage original sur l'aversion à l'ambiguïté. Dans de nombreux cadres expérimentaux, les sujets témoignent d'une forte préférence pour les enjeux risqués (i.e. ceux pour lesquels les probabilités sont connues) par rapport aux enjeux ambigus (i.e. ceux pour lesquels les probabilités sont inconnues). Bien que la théorie de la décision traite l'ambiguïté comme un simple cas particulier de la prise de décision risquée, plusieurs études de neuroéconomie montrent que les supports des mécanismes neuronaux des décisions risquées et ambiguës sont en réalité distincts. On manifeste une aversion à l'ambiguïté parce que l'on anticipe une réaction émotionnelle négative plus intense en face d'une situation ambiguë plutôt que risquée (Huettel *et al.*, 2006 ; Taya, 2012).

D'un autre côté, dans le cadre des décisions interactives, la neuroéconomie sociale aide à mieux préciser le contenu émotionnel de l'« attention portée à l'autre » comme facteur explicatif des soi-disant « anomalies » ou « irrationalités » observées : participation volontaire à des actions collectives, aide apportée à des personnes parfaitement étrangères, sur-communication, punition coûteuse envers ceux qui violent les normes sociales... Elle explicite ce faisant les bases neuronales des motivations que l'économie comportementale prête communément aux sujets pour expliquer leur comportement coopératif : confiance, altruisme, réciprocité, empathie, générosité, culpabilité, souci de l'équité, etc.

Il ne faut pas se cacher d'ailleurs que les résultats de certaines de ces expériences peuvent se révéler parfois déroutantes. Deux exemples à titre illustratif. Le premier exemple traite de l'altruisme – dont le développement est le plus souvent expliqué par un apprentissage actif ou par des influences sociales – qui passerait par des changements dans la taille et l'activation d'une certaine structure neuronale. C'est ce que prétendent Morishima *et al.* (2012), qui ont examiné les bases neuroanatomiques des différences individuelles observables en matière d'altruisme grâce à l'une des premières expériences de neuroéconomie basées sur l'utilisation d'un outil d'IRM morphologique pour scanner les cerveaux de sujets dont les paramètres de préférences sociales étaient estimés sur un jeu du dictateur et un jeu de réciprocité.

Ils établissent que le volume de la matière grise de la *jonction temporo-pariétale*³⁷ droite est fortement corrélée à la fois à l'altruisme individuel et aux conditions dans lesquelles, pour l'individu en question, cette région neuronale est sollicitée durant une décision de nature altruiste. Ce qui tendrait à signifier que les différences individuelles au niveau du volume de la jonction temporo-pariétale non seulement se traduisent en différences individuelles au niveau de la propension générale à l'altruisme mais qu'en plus, elles sont responsables d'une association entre structure et fonction cérébrale en indiquant les conditions d'activation de cette région lorsque les individus ont à arbitrer entre des comportements altruistes ou égoïstes. Ce résultat mérite bien sûr d'être confirmé par des études supplémentaires. Il peut être cependant mis en relation avec certaines expériences de neuroscience cognitive qui établissent qu'il est possible d'observer des modifications cérébrales structurelles liées à un apprentissage prolongé. Par exemple, après un apprentissage pendant plusieurs jours d'un comportement moteur manuel complexe, on observe une augmentation de la taille des zones motrices de représentation de chacun des doigts (Dayan et Cohen, 2011). De même, on mesure de plus forts volumes de cortex moteur et auditif chez les musiciens par comparaison des mêmes zones neuronales chez des non-musiciens (Münste, Altenmüller et Jäncke, 2002).

Le deuxième exemple est relatif à l'honnêteté ou l'aptitude spontanée à dire la vérité ou à mentir, une spécificité individuelle de nature profondément morale, qui selon Karton et Bachmann (2011) serait liée entre autres à l'activation d'une certaine zone cérébrale. En combinant la stimulation magnétique transcrânienne et l'imagerie cérébrale par IRMf, les auteurs établissent que la propension spontanée au mensonge peut être manipulée en exerçant une stimulation au niveau du cortex préfrontal dorsolatéral. Ils emploient dans l'expérience une tâche, sans pénalité, de reconnaissance visuelle d'objets. Ils montrent, plus précisément, que la stimulation de cette zone dans l'hémisphère droit tend à freiner la propension au mensonge tandis que dans l'hémisphère gauche elle tend au contraire à la favoriser. Ainsi, notre sens du bien et du mal ne serait pas seulement fonction de l'éducation, des croyances religieuses et philosophiques, mais aussi de la biologie de notre cerveau... Là encore, les résultats de cette étude exploratoire gagneraient bien sûr à être confirmés, en particulier avec des paradigmes expérimentaux plus réalistes qui introduiraient certaines raisons de dire la vérité ou mentir³⁸.

D'un autre côté, certaines expériences de neuroéconomie témoignent des contraintes que font peser les spécificités de l'activité cérébrale sur le raisonnement stratégique des individus. Ainsi est-il établi notamment que le cerveau humain est incapable de dépasser deux à trois niveaux d'itérations dans les processus de raisonnement usuels : le fonctionnement même du cerveau limite la « profondeur de raisonnement ». L'expérience menée par Coricelli et Nagel (2009), qui prend le jeu du « concours de beauté » comme paradigme comportemental, illustre pleinement ce résultat³⁹.

³⁷ Les aires de la jonction temporo-pariétale, à la lisière des lobes temporaux et pariétaux, interviennent dans le processus de mentalisation (*mentalizing*) en même temps que dans le jugement moral. La mentalisation, qualifiée aussi de théorie de l'esprit (*theory of mind*), est un concept élaboré à l'origine en psychologie sociale. Il se réfère à la capacité cognitive des humains de distinguer leurs propres états mentaux (croyances, intentions, désirs...) des états mentaux d'autrui afin d'utiliser cette distinction pour anticiper le comportement des autres et guider en retour leur propre comportement.

³⁸ On trouve aujourd'hui dans la littérature expérimentale et neuroéconomique plusieurs études consacrées à l'honnêteté et au mensonge (voir notamment Wibral *et al.*, 2012 ; Baumgartner, Gianotti et Knoch, 2013 ; Abe et Greene, 2014 ; Fioretti et Marden, 2015). Ces travaux renvoient au développement d'une littérature sur le « neurodroit », en liaison avec le recours croissant à l'imagerie fonctionnelle aux Etats-Unis dans le contexte des procès criminels, et à l'apparition d'un « marché de la vérité », avec la création de puissantes sociétés (comme *No Lie MFR* et *Cephos*) qui proposent leurs services dans la neurodétection du mensonge (pour des discussions sur le neurodroit et les problèmes éthiques que cette nouvelle démarche soulève, voir notamment Sinnott-Armstrong *et al.*, 2008 ; Oullier, 2012 ; Forest, 2014, chap. 1).

³⁹ Cette expérience de neuroéconomie est à mettre en parallèle avec un certain nombre d'expériences de psychologie cognitive plus anciennes qui témoignent des limites cognitives du cerveau. Par exemple, dans Miller (1956), il est établi que le cerveau humain ne peut traiter au mieux que sept données à la fois. Le concept de « comptabilité mentale », introduit en économie comportementale pour rendre compte des

Les expériences neuroéconomiques traitent de questions économiques variées : décisions de consommation courantes, choix financiers (pour lesquelles l'attitude à l'égard du risque et de l'incertitude tiennent une place prépondérante), choix intertemporels (pour lesquels l'escompte psychologique du futur et l'impulsivité entrent en jeu), choix interpersonnels et décisions sociales (pour lesquelles interviennent des questions de cognition sociale en même temps que l'influence des normes sociales/morales/culturelles dans la relation à l'autre). En dépit des fortes spécificités de chacune de ces classes de problème de décision, certaines régularités qualitatives dans les résultats des expériences peuvent être d'ores et déjà identifiées. Dans le champ des décisions interactives comme dans celui des décisions individuelles, il est désormais acquis que les émotions et les affects participent pleinement à la prise de décision économique, en liaison étroite avec les processus cognitifs plus élaborés de délibération et d'évaluation. C'est là un enseignement unanimement reconnu. Pour Sanfey *et al.* (2006), (Bourgeois-) Gironde (2008) ou Oullier (2010), par exemple, la mise en évidence du rôle des émotions dans la prise de décision économique constitue le point commun de l'ensemble des travaux de neuroéconomie.

Cet enseignement général, à vrai dire, se décline diversement : tantôt, les émotions sont présentées comme indispensables à une prise de décision « rationnelle » (dans le prolongement des découvertes menées par le neuropsychologue Antonio Damasio et son équipe dans les années 1990 : Bechara *et al.*, 1994, 1997), tantôt elles sont censées intervenir en concertation étroite avec les processus cognitifs de contrôle et d'évaluation (ce qui permet d'expliquer en particulier certains comportements apparemment irrationnels, en conformité avec certains modèles d'économie comportementale), tantôt encore les émotions semblent agir en concurrence avec les traitements cognitifs de la décision, le cerveau arbitrant en quelque sorte entre raison et émotions⁴⁰. En outre, s'agissant des expériences sur les choix interpersonnels destinées notamment à expliquer la nature pro-sociale de certains comportements observés, elles confirment que les systèmes de la cognition sociale sont bien présents dans les mécanismes cérébraux en œuvre, tout particulièrement, le phénomène de la mentalisation (*mentalizing*) et le système des neurones miroirs, comme composantes de l'empathie (voir notamment Rizzolatti et Sinigaglia, 2011 ; Singer et Tusche, 2014 ; Serra, 2016, chap. 2 ; Alos-Ferrer, 2018).

D'une manière générale, il est indéniable que c'est lorsque les expériences neuroéconomiques ne se préoccupent pas seulement de la seule *localisation* des aires cérébrales activées en même temps que les sujets se livrent à certaines tâches comportementales mais qu'elles s'intéressent aussi au *fonctionnement* de l'activité cérébrale que leur apport se révèle plus instructif pour l'économiste. C'est lorsqu'elles ne se contentent pas de chercher « où » sont encodés dans le cerveau les signaux associés aux décisions économiques mais également « comment » et « quand » que s'ouvrent des perspectives plus prometteuses pour l'économiste (Billot, Bourgeois-Gironde et Corcos, 2012 ; Hsu et Zhu, 2012 ; Taya, 2012). À ce titre, la plupart des critiques émises à l'encontre de la neuroéconomie visent en réalité la seule économie comportementale dans le scanner, voire simplement pour l'essentiel les toutes premières de ces expériences en neuroéconomie sociale. S'agissant des travaux participant du programme de l'économie neuronale, centrés à l'origine sur les problèmes séquentiels d'apprentissage de la récompense traités par le système nerveux, ils échappent à la critique.

simplifications mentales opérées par le cerveau pour traiter des choix complexes (voir notamment Thaler, 1985, 1999) trouve ainsi une justification objective dans cette capacité limitée de traitement de l'information par le cortex préfrontal.

⁴⁰ La plupart des résultats de l'économie comportementale dans le scanner est souvent interprétée, encore aujourd'hui, dans une seule « perspective dualiste », opposant des choix « instinctifs, émotionnels » à des choix « raisonnés, cognitifs » (Vallois, 2012, par exemple). La grille de lecture moins radicale retenue ici paraît plus conforme à la diversité des résultats auxquels parviennent ces travaux (voir Serra 2016, chap. 4).

L'économie neuronale

Rappelons brièvement qu'on admet aujourd'hui en neuroscience que le système cérébral de récompense est indispensable à la survie, car il fournit la motivation nécessaire à la réalisation de comportements adaptés, permettant de préserver l'individu et l'espèce (recherche de nourriture, reproduction, évitement des dangers...). Si, dans le règne animal, la survie et la reproduction conditionnent l'essentiel des récompenses, chez l'homme moderne ce modèle purement physiologique de menace pour l'espèce n'apparaît plus que de manière secondaire ; la menace apparaît lors de situations de contrainte plus ou moins subjectives, comme par exemple lors de prise de décisions complexes. Cette conception du système de récompense reflète un changement de paradigme radical intervenu en neuroscience dans les années 1990 : les actions ne sont plus censées régies uniquement par des réflexes de type pavlovien (travaux de Sherrington datant du tout début du 20^{ème} siècle) mais aussi et surtout par des motivations et des intentions (Glimcher, 2003). Les systèmes « réflexes » sont conçus comme une simple classe particulière de mécanismes d'apprentissage parmi d'autres qui valorisent seulement un petit ensemble de comportements solidement ancrés au plan de l'évolution. Deux autres systèmes au moins co-existent dans le cerveau : les systèmes « routiniers » (*habitual systems*), capables d'apprendre à valoriser un large ensemble d'actions pas un processus simple d'essai-erreur mais s'appliquant seulement dans des environnements suffisamment stables ou se transformant selon un rythme lent, et les systèmes « téléologiques » (*goal-oriented systems*), qui valorisent les actions en fonction de leur impact anticipé et sont capables de calculer la valeur de chaque action dans des situations entièrement nouvelles et dans des environnements en rapide évolution (Bouton, 2007 ; Rangel, Camerer et Montague, 2008 ; Balleine, Daw et O'Doherty, 2009 ; Daw et O'Doherty, 2014 ; Rolls, 2014). Avec l'apparition dans les années 1990 de ce nouvel outil expérimental que constitue l'électrophysiologie, les expériences permettent d'enregistrer l'activité individuelle des neurones d'animaux conscients avant qu'ils n'effectuent leurs choix. Dans la mesure où l'observation précède la décision, cela rend accessible la compréhension du mécanisme en jeu : on devient capable de comprendre comment le singe, conditionné à effectuer des mouvements oculaires, forme et attribue une valeur subjective à chaque option qui lui est soumise (Schultz, Dayan et Montague, 1997).

Grâce à la neuro-imagerie, au tournant des années 2000, le paradigme de l'apprentissage de la récompense a été étendu de l'animal à l'homme dans une série d'expériences faisant intervenir des récompenses gustatives ou financières (Delgado *et al.*, 2000 ; Knutson *et al.*, 2000 ; Elliot *et al.*, 2000 ; Knutson *et al.*, 2001 ; Berns *et al.*, 2001 ; Breiter *et al.*, 2001) : même si l'on observe certaines différences cytoarchitectoniques entre le cerveau du singe et celui de l'homme (Mackey et Petrides, 2010, 2014 ; Wallis, 2012), les *processus cérébraux* semblent similaires. Ces travaux montrent que, chez l'homme comme chez les primates non-humains, la *stochasticité* et la *séquentialité* apparaissent comme deux propriétés saillantes du mécanisme de choix au plan neuronal. Mais la question de la « valeur subjective » d'une décision est tout autant centrale pour comprendre le fonctionnement du cerveau. L'un des faits marquants de l'économie neuronale est au fond d'avoir transféré le concept de valeur de la sphère économique à la sphère neurologique.

On peut considérer, de manière très schématique, qu'au programme de l'économie neuronale sont associées au moins trois idées majeures : (i) le cerveau « valorise » en termes de récompenses/punitions de multiples stimuli, et ceci indépendamment des processus sensoriels ou moteurs (Platt et Glimcher, 1999 ; Padoa-Schioppa et Assad, 2006 ; Plassman, O'Doherty et Rangel, 2007) ; (ii) il existe un « numéraire commun » (*common currency*) qui permet au cerveau de comparer sur une même échelle les valeurs subjectives de ces multiples stimuli *a priori* incomparables ; la production de dopamine (l'un des principaux neurotransmetteurs qui envoie au plan chimique les signaux d'un neurone à l'autre) apparaît à

cet égard comme l'une des composantes majeures du mécanisme neuronal *centralisé* qui résoudrait le problème de la comparabilité (Montague et Berns, 2002 ; Landreh et Bickle, 2008 ; Schultz, 2009 ; Balleine, Daw et O'Doherty, 2009 ; Niv et Montague, 2009 ; Rangel, 2009 ; Chib *et al.*, 2009 ; Levy et Glimcher, 2012) ; (iii) l'incertitude constitue une donnée fondamentale pour comprendre le fonctionnement du cerveau (Glimcher, 2003 ; Glimcher, Dorris et Bayer, 2005 ; Faisal, Selen et Wolpert, 2008 ; Rolls et Deco, 2010 ; Glimcher, 2011a ; Fehr et Rangel, 2011 ; Rolls, 2014) ; confronté à cette incertitude, le cerveau aurait la capacité de faire des hypothèses et d'anticiper, consolidant ou modifiant ses croyances après vérifications ; il ferait en permanence des « mises à jour » via divers *mécanismes d'apprentissage* que l'on commence à identifier (Niv et Montague, 2009 ; Balleine et O'Doherty, 2010 ; Daw et Tobler, 2014)⁴¹. Ce qu'enseigne fondamentalement l'économie neuronale c'est que, dans la compréhension des mécanismes neuronaux qui régissent les décisions, ces trois propriétés interdépendantes sont essentielles.

Plutôt que de se contenter d'emprunter leurs modélisations à l'économie comportementale, ces travaux menés en économie neuronale visent ainsi à élaborer une véritable « théorie neuronale des comportements d'apprentissage ». Ce faisant, la neuroéconomie prend ses distances avec l'économie comportementale. Car les « anomalies » ou « biais », en particulier, n'y ont plus la même signification : ce n'est plus par référence à une certaine norme économique-socio-culturelle que certains comportements peuvent être disqualifiés comme « irrationnels » mais par rapport à un processus d'apprentissage neuronal « optimal »⁴².

Il faut enfin relever que l'orientation récente des recherches en neuroéconomie – que l'on peut interpréter comme l'émergence d'une « économie neuronale et comportementale » – conforte la spécificité de la neuroéconomie dans sa capacité à proposer aux économistes des enseignements inédits (Vromen, 2011 ; Serra, 2016, chap. 4).

L'économie neuronale et comportementale

L'ambition commune de nombreux neuroéconomistes est aujourd'hui d'élaborer un « modèle neuroéconomique de choix » qui intégrerait ces contraintes que le fonctionnement cérébral fait peser sur la prise de décision économique, un modèle dont les premiers contours explicitent d'ores et déjà sur une base nouvelle les faiblesses de la théorie économique standard.

⁴¹ Fondamentalement, la nature aléatoire des signaux neuronaux provient de cette nécessité dans laquelle le cerveau se trouve d'élaborer des stratégies d'apprentissage des récompenses. La prédiction des récompenses est conçue pour s'appliquer à des choix entre options dont les probabilités de gains associées sont au départ inconnues mais apprises progressivement. La propriété centrale est que le signal neuronal ne porte pas sur des *choix isolés* mais sur des *séquences de choix répétés*. Il peut prendre ainsi des valeurs variables lors des répétitions quand bien même le choix serait resté parfaitement identique (Glimcher, 2003). Pour l'économiste, cette propriété stochastique du mécanisme de choix neuronal est tout autant primordiale que l'hypothèse du « numéraire commun ». Car elle indique en particulier qu'il existe une différence notable entre le traitement de l'incertain par le cerveau et la manière dont les économistes abordent ordinairement la question du risque. Quasiment tous les modèles économiques, dans les situations risquées, effectuent une séparation entre d'un côté les probabilités d'occurrence des divers états possibles et, d'un autre côté, les utilités attendues en ces états. Et ces probabilités et ces préférences sont supposées données. En neurobiologie cette séparation ne tient plus : le fonctionnement du cerveau suggère plutôt une séparation entre d'un côté, l'encodage des *renforcements anticipés*, et de l'autre celui du *risque* qui entache forcément le processus de choix, dans la mesure où la nature des renforcements est nécessairement stochastique. La dynamique d'apprentissage porte à la fois sur le montant de la récompense et sa probabilité d'occurrence, qui représentent deux dimensions inséparables du choix (Bosschaerts, Preuschoff et Hsu, 2009, notamment). L'incertitude est présente tout au long du système de prise de décision. C'est là une contribution majeure des neurosciences à la compréhension des mécanismes de la décision : il n'y a pas de choix sans risques. Cela implique en particulier que les renforcements identifiés en neurosciences ne doivent pas être confondus avec les utilités espérées que calculent classiquement les économistes – confusion que la qualification d'« utilité espérée physiologique » proposée un temps par Glimcher pouvait entretenir (Glimcher, Dorris et Bayer, 2005). D'autant qu'une partie des renforcements n'est pas sous le contrôle des fonctions cognitives supérieures mais est générée par le système des émotions largement involontaire. Le fonctionnement du cerveau en termes de récompenses/punitions fait passer au second plan la distinction entre systèmes « rationnels » ou « émotionnels ».

⁴² Une interprétation intéressante est avancée par Vallois (2011) : dans ces travaux d'économie neuronale, ce seraient les cas pathologiques qui serviraient de substitut à l'absence de norme.

Ce modèle présente la particularité d'intégrer les nombreux résultats expérimentaux accumulés depuis une quinzaine d'années par l'économie comportementale dans le scanner au cadre théorique développé par l'économie neuronale. Schématiquement, deux mécanismes sont supposés intervenir de manière séquentielle au plan cérébral lors de toute décision : (i) un mécanisme de *valorisation*, qui est censé refléter la « valeur » attachée par le cerveau aux différentes options en jeu dans le problème de choix ; (ii) un mécanisme de *sélection*, qui s'appuyant sur les résultats du mécanisme de valorisation est censé générer par comparaison le choix effectif de l'une des options⁴³. Figure ainsi au centre du dispositif le concept de « valeur subjective » d'une décision, qui est mis en relation avec la notion d'utilité « ressentie » (ou « expérimentée ») (*experienced utility*). Si la valeur d'une décision correspond à un signal calculé au moment du choix, l'utilité « ressentie » est censée représenter quant à elle un signal calculé au moment de l'expérience effective par l'organisme des conséquences de l'option choisie (Kable et Glimcher, 2009 ; Rangel et Hare, 2010 ; Fehr et Rangel, 2011 ; Padoa-Schioppa, 2011 ; Rangel et Clithero, 2014 ; Berridge et O'Doherty, 2014 ; Glimcher, 2011a, 2014). On ne peut manquer bien sûr de tracer un parallèle entre cette distinction et celle, souvent avancée en économie comportementale, entre l'utilité « décisionnelle », qui correspond au concept d'utilité de la théorie économique standard et l'utilité « expérimentée », qui possède un caractère hédoniste et symbolise le bien-être subjectif (Kahneman *et al.*, 1997 ; Kahneman, 2003, 2012, notamment).

Les principaux résultats à ce jour établis peuvent se résumer ainsi : i) les taux d'activation des neurones encodent des valeurs *relatives* (et non absolues) (*i.e.* elles dépendent de la valorisation des autres options disponibles et en particulier d'un point de référence lié au contexte de la décision) ; ii) l'activité neuronale est intrinsèquement *stochastique* ; confronté à cette incertitude, le cerveau a la capacité de faire des hypothèses et d'anticiper, consolidant ou modifiant ses croyances après vérifications ; il fait en permanence des « mises à jour » via divers mécanismes d'apprentissage (hypothèse du « cerveau bayésien ») ; iii) au niveau de l'implémentation du choix par le cerveau, l'algorithme du « vainqueur empoche le tout » (*i.e.* l'équivalent neuronal de l'opération « arg max » du modèle économique standard : la maximisation de l'utilité) semble supplanté par l'algorithme plus direct de la « course au seuil » (*drift-diffusion*) : une décision est prise dès lors que les taux d'activation des neurones correspondants excèdent un certain seuil critique⁴⁴. Ce que nous apprend le modèle neuroéconomique de choix, fondamentalement, c'est que l'erreur est consubstantielle aux choix économiques : la décision « optimale » est celle qui correspond à l'option associée au signal d'utilité ressentie le plus élevé au moment de la « consommation » ; si ce n'est pas le cas, l'individu commet une erreur de décision. En outre, « le modèle ne se contente pas de signaler que les erreurs sont probables mais il éclaire la manière dont des variables économiques, comme le coût d'opportunité du temps ou les interventions de marketing, sont susceptibles d'affecter leur occurrence » (Fehr et Rangel, 2011, p. 15)⁴⁵.

⁴³ A noter qu'au sein de ces deux mécanismes, le niveau d'attention des sujets tient un rôle central comme modulateur des signaux.

⁴⁴ Comment ce seuil est-il déterminé ? Cela dépendrait de la nature plus ou moins complexe des décisions. Le cerveau effectuerait un arbitrage entre vitesse et exactitude : si le seuil est bas, le choix est grandement simplifié, au risque toutefois de ne pas choisir la meilleure option, alors que s'il est élevé, le choix devient plus difficile et plus long, mais le risque d'erreur a tendance à décroître (Roitman et Shadlen, 2002 ; Palmer, Huk et Shadlen, 2005). La littérature sur les modèles de course au seuil est vaste. Voir notamment la synthèse proposée par Ratcliff *et al.* (2016).

⁴⁵ La neuroéconomie, à cet égard, est présentée par certains auteurs comme un instrument d'orientation potentiel du choix des consommateurs. La neuroéconomie fournirait une base neurobiologique à la politique de marketing engagée par les entreprises afin d'orienter les choix des consommateurs à leur insu sur tel ou tel produit (changer l'attractivité visuelle du produit et de son emballage, par exemple). De fait, le « neuromarketing » fait débat sur un plan éthique : s'agit-il d'« une manipulation mentale » des consommateurs, plus poussée que les publicités conventionnelles, qui amputerait leur liberté de choix (Courbet et Benoit, 2013, par exemple) ou d'un simple instrument de connaissance des mécanismes neuronaux en jeu lors des décisions d'achat dont les enseignements peuvent très bien être mis au service

Bien sûr, ce modèle reste encore largement exploratoire. Il s'applique essentiellement aux décisions téléologiques (*goal-directed decisions*) et il ne concerne que les choix « simples » (*i.e.* portant sur un nombre réduit de biens courants, sans asymétries informationnelles, considérations stratégiques, problèmes de maîtrise de soi, attention à l'autre, dimension financière ou temporelle). Les auteurs espèrent que ce modèle puisse fournir un banc d'essai satisfaisant pour l'étude systématique des questions économiques et que certaines de ses propriétés essentielles en termes calculatoire et neurobiologique aient toutes les chances d'être préservées dans le cas de choix plus complexes. Comme semblent le suggérer d'ailleurs les premiers résultats disponibles portant sur les choix risqués (Levy *et al.*, 2010), les choix intertemporels (Kable et Glimcher, 2009 ; Fehr et Rangel, 2011 ; Kable, 2014) ou les décisions sociales (Hare *et al.*, 2010 ; Ruff et Fehr, 2014 ; Hutcherson, Bushong et Rangel, 2015 ; Krajbich *et al.*, 2015), cette hypothèse paraît raisonnable.

Remarque finale

La neuroéconomie est un champ disciplinaire tout récent ; deux décennies au plus. Comme toute discipline émergente, elle soulève bon nombre d'interrogations que l'article a regroupé autour de trois questions majeures portant sur le statut des connaissances à ce jour produites : des connaissances certaines ? des connaissances scientifiquement justifiées ? des connaissances utiles à l'économiste ? Sans prétendre à l'exhaustivité, ce cadre d'analyse permet d'apprécier les lignes de force des principales critiques adressées à la neuroéconomie en même temps que la nature des paradigmes susceptibles d'être déployés.

L'article ne rend pas compte en revanche de l'ensemble des réflexions critiques « externes » adressées plus généralement aux neurosciences modernes (et par ricochet à la neuroéconomie), des réflexions visant plus particulièrement la prétention des neurosciences à identifier connaissance du cerveau, connaissance de soi et connaissance de la société, qui sont développées avec plus ou moins de véhémence dans une littérature anthropo-socio-philosophique (voir notamment Ehrenberg, 2008, 2015).

Références bibliographiques

- Abe N. and J.D. Greene (2014), "Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty", *Journal of Neuroscience*, 34(32): 10564-10572.
- Allos-Ferrer C. (2018), "A review essay on *Social neuroscience: Can research on the social brain and economics inform each other?*" *Journal of Economic Literature*, 56(1): 234-264.
- Aydinonat N.M. (2010), "Neuroeconomics: more than inspiration, less than revolution", *Journal of Economic Methodology*, 17(2): 159-169.
- Aydogan G., N.C. Furtner, B. Kern, A. Jobst, N. Müller, and M.G. Kocher (2017), "Oxytocin promotes altruistic punishment", *Social Cognitive and Affective Neuroscience*, 1540-1747.
- Baker L. (1989), "Instrumental intentionality", *Philosophy of Science*, 56: 303-316.
- Balleine B.W, N.D. Daw et J.P. O'Doherty (2009), "Multiple forms of value learning and the function of dopamine", in Glimcher P. *et al.* (eds.) (2009), chap.: 24: 367-388.
- Balleine B.W and J.P. O'Doherty (2010), "Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action", *Neuropsychopharmacology*, 35: 48-69.
- Bardsley N., R. Cubitt, G. Loomes, P. Moffatt, and R. Sugden (2010), *Experimental economics: rethinking the rules*, Princeton: Princeton University Press.
- Barraza J.A., M.E. McCulloch, S. Ahmadi, and P.J. Zak (2011), "Oxytocin infusion increases charitable donations regardless of monetary resources", *Hormones and Behavior*, 60(2): 148-151.

d'autres agents de la vie économique que les entreprises, comme par exemple les associations de consommateurs ((Bourgeois-) Gironde, 2008 ; Droulers et Roule., 2010) ?

- Barraza J.A. and P.J. Zak (2009), "Empathy toward stranger triggers oxytocin release and subsequent generosity", *Annals of the New York Academy of Science*, 1167: 182-189.
- Baumgartner T., L.R.R. Gianotti, and D. Knoch (2013), "Who is honest and why: baseline activation in anterior insula predicts inter-individual differences in deceptive behavior", *Biological Psychology*, 94: 192-197.
- Baumgartner T., M. Heinrichs, A. Vonlanthen, U. Fischbacher, and E. Fehr (2008), "Oxytocin shapes the neural circuitry of trust and trust adaptation in humans", *Neuron*, 58: 639-650.
- Baumgartner T., T. Knoch, D. Hotz, P. Eisenegger, and C. Fehr (2011), "Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice", *Nature Neuroscience*, 14(11): 1468-1474.
- Bechara A., A.R. Damasio, H. Damasio, and S.W. Anderson (1994), "Insensitivity to future consequences following damage to human prefrontal cortex", *Cognition*, 50: 7-15.
- Bechara A., H. Damasio, D. Tranel, and A.R. Damasio (1997), "Deciding advantageously before knowing the advantageous strategy", *Science*, 275 (February 28): 1293-1295.
- Becker G. (1976), "Altruism, egoism, and genetic fitness: economics and sociobiology", *Journal of Economic Literature*, 14(3): 817-826.
- Bernheim B.D. (2009), "On the potential of neuroeconomics: a critical (but hopeful) appraisal", *American Economic Journal: Microeconomics*, 1(2): 1-41.
- Berns G.S., S.M. McClure, G. Pagnoni, and P.R. Montague (2001), "Predictability modulates human brain responses to reward", *Journal of Neurosciences*, 21: 2793-2789.
- Berridge K.C. and J.P. O'Doherty (2014), "From experienced utility to decision utility", in Glimcher P. and E. Fehr (eds.) (2014), chap.17: 335-351.
- Billot A., S. Bourgeois-Gironde, and A. Corcos (2012), "Introduction: neuroeconomics of learning and decision", *Recherches économiques de Louvain*, 78(3): 5-11.
- Bird A. (1998), *Philosophy of science*, Montreal & Kingston: McGill-Queen's University Press.
- Bossaerts P., K. Preuschoff, and M. Hsu (2009), "The neurobiological foundations of valuation in human decision making under uncertainty", in Glimcher P. et al. (eds.) (2009), chap. 23: 353-366.
- (Bourgeois-)Gironde S. (2008), *La neuroéconomie. Comment le cerveau gère mes intérêts*, Paris : Plon.
- Bourgeois-Gironde S. (2010), "Is neuroeconomics doomed by the reverse inference fallacy?" *Mind & Society*, 9(2): 229-249.
- Bouton M.E. (2007), *Learning and behaviour: a contemporary synthesis*, Sinauer Associates Inc., Sunderland, Massachusetts.
- Breiter H.C., I. Aharon, D. Kahneman, A. Dale, and P. Shizgal (2001), "Functional imaging of neural responses to expectancy and experience of monetary gains and losses", *Neuron*, 30: 619-639.
- Camerer C.F. (2003), *Behavioral game theory. Experiments in strategic interaction*, Princeton: Princeton University Press.
- Camerer C.F. (2006), "Behavioral economics", Working Paper Caltech, Div HSS 228-77 (prepared for the World Congress of the Econometric Society, august 2005, London).
- Camerer C.F. (2007), "Neuroeconomics: using neuroscience to make economic predictions", *Economic Journal*, 117: C26-C42.
- Camerer C.F. (2008a), "The case of mindful economics", in Caplin A. and A. Schotter (eds.) (2008), chap. 2: 43-69.
- Camerer C.F. (2008b), "The potential of neuroeconomics", *Economics and Philosophy*, 24: 369-379.
- Camerer C.F., G. Loewenstein, and D. Prelec (2004), "Neuroeconomics: why economics needs brains", *Scandinavian Journal of Economics*, 106(3): 555-579.
- Camerer C.F., G. Loewenstein, and D. Prelec (2005), "Neuroeconomics: how neuroscience can inform economics", *Journal of Economic Literature*, 43: 9-64.
- Camerer C.F., G. Loewenstein, and M. Rabin (eds.) (2004), *Advances in behavioural economics*, Princeton, NJ: Russell Sage Foundation, Princeton University Press.
- Camille N., G. Coricelli, J. Sallet, P. Pradat-Diehl, J-R. Duhamel, and A. Sirigu (2004), "The involvement of the orbitofrontal cortex in the experience of regret", *Science*, 304, 21 may : 1167-1170.
- Camus M., S. Shimojo, C.F. Camerer, J. O'Doherty, and A. Rangel (2009), "rTMS of right dorsolateral prefrontal cortex disrupts the computation of goal-directed values", *European Journal of Neuroscience*, 30: 1980-1988.

- Caplin A. and M. Dean (2009), "Axiomatic neuroeconomics", in Glimcher P. *et al.* (eds.) (2009), chap. 3: 21-32.
- Caplin A. and P. Glimcher (2014), "Basic methods from neoclassical economics", in Glimcher P. and E. Fehr (eds.) (2014), chap. 1: 3-17.
- Caplin A. and A. Schotter (eds.) (2008), *The foundation of positive and normative economics: a handbook*, New York: Oxford University Press.
- Chalmers A.F. (1999), *What is this thing called science?* Maidenhead and New York: Open University Press.
- Chib V.S., A. Rangel, S. Shimojo, and J.P. O'Doherty (2009), "Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex", *Journal of Neuroscience*, 29: 12315-12320.
- Churchland P.S. (2011), *Braintrust: what neuroscience tells us about morality*, Princeton, NJ: Princeton University Press.
- Clément V., C. Le Clairche et D. Serra (2008), *Economie de la justice et de l'équité*, Paris : Economica.
- Coates J.M. and J. Herbert (2008), "Endogenous steroids and financial risk taking on a London trading floor", *Proceedings of the National Academy of Sciences U.S.A.*, 105: 6167-6172.
- Coricelli G., H.D. Critchley, M. Joffily, J.P. O'Doherty, A. Sirigu, and R. J. Dolan (2005), "Regret and its avoidance: a neuroimaging study of choice behaviour", *Nature Neuroscience*, 8(9): 1255-1262.
- Coricelli G. and R. Nagel (2009), "Neural correlates of depth of strategic reasoning in medial prefrontal cortex", *Proceedings of the National Academy of Science USA*, 106: 9162-9168.
- Corrado G.S. and K. Doya (2007), "Understanding neural coding through the model-based analysis of decision making", *Journal of Neuroscience*, 27: 8178-8180.
- Corrado G.S., L.P. Sugrue, J.R. Brown, and W.T. Newsome (2009), "The trouble with choice: studying decision variables in the brain", in Glimcher P. *et al.* (eds.) (2009), chap. 30: 463-480.
- Cot A.L. et S. Ferrey (2016), « La construction de 'faits' économiques d'un nouveau type : éléments pour une histoire de l'économie expérimentale », *L'Actualité économique. Revue d'analyse économique*, 92(1-2) : 11-47.
- Courbet D. et D. Benoit (2013), « Neurosciences au service de la communication commerciale : Manipulation et éthique – Une critique du neuromarketing », *Etudes de communication*, 40 : 28-42.
- Cozic M. (2012), "Economie 'sans esprit' et données cognitives", *Revue de philosophie économique*, 13 (1) : 127-153 (numéro coordonné par D. Serra : Experimental economics : some methodological aspects).
- Craig A.D. (2009), "How do you feel now? The anterior insula and human awareness", *Nature Reviews Neuroscience*, 10: 59-70.
- Cravel C.F. (2007), *Explaining the brain: mechanisms and the mosaic unity of neuroscience*, Oxford: Oxford University Press.
- Crockett M.J., L. Clark, G. Tabibnia, M.D. Liebermann, and T.W. Robbins (2008), "Serotonin modulates behavioural reactions to unfairness", *Science*, 320: 1739.
- Crockett M.J. and E. Fehr (2014), "Pharmacology of economic and social decision making", in Glimcher P. and E. Fehr (eds.) (2014), chap. 14: 259-279.
- Daw N.D. and J.P. O'Doherty (2014), "Multiple systems for value learning", in Glimcher P. and E. Fehr (eds.) (2014), chap. 21: 393-410.
- Daw N.D. and P.N. Tobler (2014), "Value learning through reinforcement: the basics of dopamine and reinforcement learning", in Glimcher P. and E. Fehr (eds.) (2014), chap. 15: 283-298.
- Dawson M.E., A.M. Schell, and C.G. Courtney (2011), "The skin conductance response, anticipation, and decision-making", *Journal of Neuroscience, Psychology, and Economics*, 4(2): 111-116.
- Dayan E. and L.G. Cohen (2011), "Neuroplasticity subserving motor skill learning", *Neuron*, 72: 443-454.
- Declerck C. et C. Boone (2016), "The neuroanatomy of prosocial decision making" in Declerck C. and C. Boone (eds.), *Neuroeconomics of prosocial behaviour. The compassionate egoist*, London: Elsevier.
- De Dreu K. (2012a), "Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion", *Psychoneuroendocrinology*, 37: 871-880.
- De Dreu K. (2012b), "Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda", *Hormones and Behavior*, 61: 419-428.
- Delgado M.R., L.E. Nystrom, C. Fissel, D.C. Noll, and J.A. Fiez (2000), "Tracking the hemodynamic responses to reward and punishment in the striatum", *Journal of Neurophysiology*, 84: 3072-3077.
- Dennett D. (1987), *The intentional stance*, Cambridge, MA: MIT Press.

- De Quervain, U. Fishbacher, V. Treyer, M. Schellhaller, U. Schnyder, A. Buck, and E. Fehr (2004), "The neural basis of altruistic punishment", *Science*, 305: 1254-1258.
- D'Esposito M, D. Ballard, G.K. Aguirre, and E. Zarahn (1998), "Human prefrontal cortex is not specific for working memory: a functional MRI study", *NeuroImage*, 8: 274-282.
- Dreher J.C. et L. Tremblay (2017), *Decision neuroscience. An integrative perspective*, Academic Press.
- Droulers O. et B. Rouillet (2010), *Neuromarketing. Le marketing revisité par les neurosciences du consommateur*, Paris : Dunod.
- Dulleck U., A. Ristl, M. Schaffner et B. Torgler (2011), "Heart rate variability, the autonomic nervous system, and neuroeconomic experiments", *Journal of Neuroscience, Psychology, and Economics*, 4(2) : 117-124.
- Ehrenberg A. (2008), « Le cerveau "social" : chimère épistémologique et vérité sociologique », *Esprit*, 341 (janvier) : 79-103.
- Ehrenberg A. (2015), « Se définir par son cerveau. La biologie de l'esprit comme forme de vie », *Esprit* (janvier) : 68-81.
- Eickhoff S.B., D. Bzdok, A.R. Laird, F. Kurth, and P. Fox (2012), "Activation likelihood estimation meta-analysis revisited", *NeuroImage*, 59(3): 2349-2361.
- Eisenegger C., J. Haushofer, and E. Fehr (2011), "The role of testosterone in social interaction", *Trends in Cognitive Sciences*, 15(6): 263-271.
- Elliot R., Z. Agnew, and J.F. Deakin (2008), "Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in human", *European Journal of Neuroscience*, 27: 2213-2218.
- Evans S.L., O. Dal Monte, P. Noble, and B.B. Averbeck (2013), "Intranasal oxytocin effects on social cognition: a critique", *Brain Research*, Science Direct (e-pub).
- Faisal A., L. Selen, and D. Wolpert (2008), "Noise in the nervous system", *Nature Reviews. Neuroscience*, 9: 292-303.
- Fann K.T. (1970), *Peirce's theory of abduction*, The Hague: Martinus Nijhoff.
- Fehr E. (2009), "Social preferences and the brain", in Glimcher P. *et al.* (2009), chap. 15: 215-232.
- Fehr E. and C.F. Camerer (2007), "Social neuroeconomics: the neural circuitry of social preferences", *Trend in Cognitive Sciences*, 11(10): 419-427.
- Fehr E. and I. Krajbich (2014), "Social preferences and the brain", in Glimcher P. and E. Fehr (2014), chap. 11: 193-218.
- Fehr E. and A. Rangel (2011), "Neuroeconomics foundations of economic choices – Recent advances", *Journal of Economic Perspectives*, 25 (4): 3-30.
- Fioretti M. and S. Marden (2015), "Suboptimal dishonesty: rationality in the absence of strategic behaviour in honesty experiments", *Journal of Neuroscience*, 35(5): 1817-1818.
- Forest D. (2014), *Neurosepticisme*, Paris: Ithaque.
- Fox C.R. and R.A. Poldrack (2014), "Prospect theory and the brain", in Glimcher P. and E. Fehr (eds.), Appendix: 533-567.
- Frith U. and C.D. Frith (2003), "Development and neurophysiology of mentalizing", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431): 459-473.
- Fudenberg D. (2006), "Advancing beyond advances in behavioural economics", *Journal of Economic Literature*, 44: 694-711.
- Gallagher H.L., A.I. Jack, A. Roepstorff and C.D. Frith (2002), "Imaging the intentional stance in a competitive game", *NeuroImage*, 16: 814-821.
- Gallistel C.R. (2009), "The neural mechanisms that underlie decision making", in Glimcher P. *et al.* (eds.) (2009), chap. 27: 419-424.
- Gilboa, I. (2009), *Theory of decision under uncertainty*, New York: Cambridge University Press.
- Glanholt M.G and E.M. Reingold (2011), "Eye movement monitoring as a process tracing methodology in decision making research", *Journal of Neuroscience, Psychology, and Economics*, 4(2): 125-146.
- Glimcher P.W. (2003), *Decisions, uncertainty, and the brain: the science of neuroeconomics*, Cambridge, MA: MIT Press.
- Glimcher P.W. (2011), *Foundations of neuroeconomic analysis*, New York: Oxford University Press.
- Glimcher P.W. (2014), "Value-based decision making", in Glimcher P. and E. Fehr (eds.) (2014), chap. 20: 373-391.
- Glimcher P.W., C.F. Camerer, E. Fehr, and R.A. Poldrack (eds.) (2009), *Neuroeconomics. Decision making and the brain*, Amsterdam: Elsevier.
- Glimcher P.W., M.C. Dorris, and H.M. Bayer (2005), "Physiological utility theory and the neuroeconomics of choice", *Games and Economic Behavior*, 52(2): 213-256.

- Glimcher P.W. and E. Fehr (eds.) (2014), *Neuroeconomics. Decision making and the brain*, 2nd ed., Amsterdam: Elsevier.
- Goldman A.I. (1988), *Epistemology and cognition*, Harvard University Press.
- Gospic K., E. Mohlin, P. Fransson, P. Petrovic, M. Johannesson, and M. Ingvar (2011), "Limbic justice – Amygdala involvement in immediate rejection in the ultimatum game", *PLoS Biology*, 9(5): 1-8
- Grabenhorst F. and E.T. Rolls (2009), "Different representations of relative and absolute subjective value in the human brain", *NeuroImage*, 48: 258-268.
- Guala F. (2005), *The methodology of experimental economics*, New York: Cambridge University Press.
- Guala F. (2008), "Experimental Economics, History of", in Durlauf S. and L. Blume (eds.), *The New Palgrave Dictionary of Economics*, Vol. 3, London: Palgrave-MacMillan, 152-156.
- Gul F. et W. Pesendorf (2005/2008), "The case of mindless economics", in Caplin A. and A. Schotter (eds.) (2008), chap. 1: 3-39.
- Hare T.A., C.F. Camerer, D.T. Knopfle, and A. Rangel (2010), "Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition", *Journal of Neuroscience*, 30: 583-590.
- Harrison G.W. (2008a), "Neuroeconomics: a critical reconsideration", *Economics and Philosophy*, 24: 303-344 (trad. française in *Revue Française d'Economie*, 2009, XXIII (avril) : 77-137).
- Harrison G.W. (2008b), "Neuroeconomics: rejoinder", *Economics and Philosophy*, 24: 533-544.
- Harrison G.W and D. Ross (2010), "The methodology of neuroeconomics", *Journal of Economic Methodology*, 17(2): 185-196.
- Hausman D.M. (1992), *The inexact and separate science of economics*, Cambridge University Press.
- Hausman D.M. (2000), "Revealed preference, belief and game theory", *Economics and Philosophy*, 16: 99-115.
- Hausman D.M. (2008), "The mindless or mindful economics: a methodological evaluation", in Caplin A. and A. Schotter (eds.) (2008), 125-151.
- Heinrichs M., B. von Dawans, and G. Domes (2009), "Oxytocin, vasopressin, and human social behavior", *Frontiers in Neuroendocrinology*, 30: 548-557.
- Henson R. (2005), "What can functional neuroimaging tell the experimental psychologists", *Quarterly Journal of Experimental Psychology*, A 58: 193-233.
- Heukelom F. (2009), "Origin and interpretation of internal and external validity in economics", NiCE Working Paper 09-111.
- Heukelom F. (2014), *Behavioral economics. A history*, New York: Cambridge University Press.
- Hsu M., M. Bhatt, R. Adolphs, D. Tranel, and C.F. Camerer (2005), "Neural systems responding to degrees of uncertainty in human decision-making", *Science*, 310: 1680-1683.
- Hsu M. and L. Zhu (2012), "Learning in games: neural computations underlying strategic learning", *Recherches Economiques de Louvain*, 78(3): 47-72.
- Huettel S.A. (2010), "Ten challenges for decision neuroscience", *Frontiers in Neuroscience*, 4: 171-185.
- Huettel S.A., C.J. Stowe, E.M. Gordon, B.T. Warner, and M.L. Platt (2006), "Neural signatures of economic preferences for risk and ambiguity", *Neuron*, 49: 765-775.
- Hutcherson C.A., B. Bushong, and A. Rangel (2015), "A neurocomputational model of altruistic choice and its implication", *Neuron*, 87: 451-462.
- Insel T.R. (2010), "The challenge of translation in social neuroscience: a review of oxytocin, vasopressin and affiliative behaviour", *Neuron*, 65: 768-779.
- Kable J.W. (2011), "The cognitive neuroscience toolkit for the neuroeconomist: a functional overview", *Journal of Neuroscience, Psychology, and Economics*, 4(2): 63-84.
- Kable J.W. (2014), "Valuation, intertemporal choice, and self-control", in Glimcher P. and E. Fehr (eds.) (2014), chap. 10: 173-192.
- Kable J.W. and P.W. Glimcher (2009), "The neurobiology of decision: consensus and controversy", *Neuron*, 63: 733-745.
- Kahneman D. (2003), "Maps of bounded rationality: psychology for behavioral economics", *American Economic Review*, 93(5): 1449-1475.
- Kahneman D. (2012), *Système 1, Système 2. Les deux systèmes de la pensée*, Paris : Flammarion ; traduction de *Thinking, fast and slow*, London: FSG (2011).
- Kahneman D. and A. Tversky (1979), "Prospect theory: an analysis of decision under risk", *Econometrica*, 47: 313-327.

- Kahneman D., P. Wakker, and R. Sarin (1997), "Back to Bentham? Explorations of experienced utility", *Quarterly Journal of Economics*, 112: 375-405.
- Karton I., and T. Bachmann (2011), "Effect of prefrontal transcranial magnetic stimulation on spontaneous truth-telling", *Behavioral Brain Research*, 225(1): 209-214.
- Kenning P. and H. Plassmann (2005), "Neuroeconomics: an overview from an economic perspective", *Brain Research Bulletin*, 67: 343-354.
- Knoch D. et E. Fehr (2007), "Resisting the power of temptations: the right prefrontal cortex and self-control", *Annual New York Academy of Science*, 1104: 123-134.
- Knoch D., L.R. Gianotti, A. Pascual-Leone, V. Treyer, M. Regard, M. Hoffman, and P. Brugger (2006a), "Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior", *Journal of Neuroscience*, 26: 6469-6472.
- Knoch D., M.A. Nitsche, U. Fischbacher, C. Eisenegger, A. Pascual-Leone, and E. Fehr (2008), "Studying the neurobiology of social interaction with transcranial direct current stimulation – The example of punishing unfairness", *Cerebral Cortex*, 18(9): 1987-1990.
- Knoch D., A. Pascual-Leone, K. Meyer, V. Treyer, and E. Fehr (2006b), "Diminishing reciprocal fairness by disrupting the right prefrontal cortex", *Science*, 314: 829-832.
- Knoch D., F. Schneider, D. Schunk, M. Hohmann et E. Fehr (2009), "Disrupting the prefrontal cortex diminishes the human ability to build a good reputation", *Proceedings of the National Academy of Science USA*, 106(49): 20895-20899.
- Knutson B.K., C.M. Adams, G.W. Fong, and D. Hommer (2000), "Anticipation of increasing monetary reward selectively recruits nucleus accumbens", *Journal of Neuroscience*, 21(16): 1-5.
- Knutson B.K., G.W. Fong, C.M. Adams, J.L. Varer, and D. Hommer (2001), "Dissociation of reward anticipation and outcome with event related fMRI", *Neuroreport*, 12: 3883-3687.
- Koppel L., D. Andersson, I. Morrison, D. Västfjäll, and Tinghög (2017), "The (null) effect of affective touch on betrayal aversion, altruism, and risk taking", *Frontiers in Behavioral Neurosciences*, 11 (December), Art 251: 1-11.
- Kosfeld M., M. Heinrichs, P. Zak, U. Fischbacher et E. Fehr (2005), "Oxytocin increases trust in humans", *Nature*, 435: 673-676.
- Krajchik I.K.C. Arnel, B. Bartling, Y. Morishima, and E. Fehr (2015), "A common mechanism underlying food choice and social decision", *PLoS Computational Biology*, 11, e1004371.
- Lambert B., C.H. Declerck, and C. Boone (2014), "Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments", *Psychoendocrinology*, 40: 60-68.
- Landreh A. and J. Bickle (2008), "Neuroeconomics, neurophysiology and the common currency hypothesis", *Economics & Philosophy*, 24: 419-429.
- Levy D.J. and P.W. Glimcher (2012), "The root of all value: a neural common currency for choice", *Current Opinion in Neurobiology*, 22: 1027-1038.
- Levy I., J. Snell, A.J. Nelson, A. Rustichini, and P.W. Glimcher (2010), "Neural representation of subjective value under risk and ambiguity", *Journal of Neurophysiology*, 103: 1036-1047.
- Lewis C.I. (1946), *An analysis of knowledge and valuation*, La Salle: Open Court.
- Lo A.W and D.V. Repin (2002), "The psychophysiology of real-time financial risk processing", *Journal of Cognitive Neuroscience*, 14: 323-339.
- Lo A.W, D.V. Repin, and B.N. Steenbarger (2005), "Fear and greed in financial markets: a clinical study of day-traders", *American Economic Review*, 95(2): 352-359.
- Lorentz T., K. McCabe, C.F. Camerer, and P.R. Montague (2007), "Neural signature of fictive learning signals in a sequential investment task", *Proceedings of the National Academy of Science USA*, 104: 9494-9498.
- Louie K.L. and B. De Martino (2014), "The neurobiology of context-dependent valuation and choice", in Glimcher P. and E. Fehr (eds.) (2014): chap. 24: 455-476.
- MacDonald K. and T.M. MacDonald (2010), "The peptide that binds: a systematic review of oxytocin and its prosocial effects in humans", *Harvard Review of Psychiatry*, 18(1): 1-21.
- Mackey S. and M. Petrides (2010), "Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains", *European Journal of Neuroscience*, 32: 1940-1950.
- Mackey S. and M. Petrides (2014), "Architecture and morphology of the human ventromedial prefrontal cortex", *European Journal of Neuroscience*, 40(5): 2777-2796.

- Mäki U. (2003), "The methodology of positive economics' does not give us *the* methodology of positive economics", *Journal of Economic Methodology*, 10: 495-506.
- McCabe K.A. (2003), "Neuroeconomics", in Nadel L. (ed.), *Encyclopedia of cognitive science*, Nature publishing Group, New York: Macmillan Pub., 294-298.
- McClure S.M., D.I. Laibson, G. Loewenstein, and J.D. Cohen (2004), "Separate neural systems value immediate and delayed monetary rewards", *Science*, 306: 503-307.
- Mikolajczak M., J.J. Gross, A. Lane, O. Corneille, Ph. De Timary, and O. Luminet (2010), "Oxytocin makes people trusting, not gullible", *Psychological Science*, 21: 1072-1075.
- Miller G.A (1956), "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *Psychological Review*, 63: 81-97.
- Montague P.R. and G.S. Berns (2002), "Neural economics and the biological substrates of valuation", *Neuron*, 36: 265-284.
- Montague P.R., G.S. Berns, J.D. Cohen, S.M. McClure, and G. Pagnoni (2002), "Hyperscanning: simultaneous fMRI during linked social interactions", *NeuroImage*, 16: 1159-1164.
- Morishima Y., D. Schunk, A. Brunhin, C.C. Ruff, and E. Fehr (2012), "Linking brain structure and activation in temporo-parietal junction to explain the neurobiology of human altruism", *Neuron*, 75: 73-79.
- Münste T.F., E. Altenmüller, and L. Jäncke (2002), "The musician's brain as a model of neuroplasticity", *Nature Reviews Neuroscience*, 3: 473-478.
- Nave G., C. Camerer, and M. McCullugh (2015), "Does oxytocin increase trust in humans? A critical review of research", *Perspectives on Psychological Science*, 10: 772-789.
- Niv Y. and P.R. Montague (2009), "Theoretical and empirical studies of learning", in Glimcher P. *et al.* (eds.) (2009), chap. 22: 331-351.
- O'Doherty J.P., A. Hampton, and H. Kim (2007), "Model-based fMRI and its application to reward learning and decision making", *Annals of the New York Academy of Sciences*, 1104: 35-53.
- Oullier O. (2010), "The useful brain: how neuroeconomics might change our views on rationality and a couple of other things", in Michel-Kerjan E. and P. Slovic (eds.), *Irrational economist: Making decisions in a dangerous world*, Public Affairs ed.
- Oullier O. (coord.) (2012), *Le cerveau et la loi : analyse de l'émergence du neurodroit*, Document de travail 2012-7 (septembre), Paris, Centre d'Analyse Stratégique.
- Padoa-Schioppa C. (2011), "Neurobiology of economic choice: a good-based model", *Annual Review of Neuroscience*, 34: 333-359.
- Padoa-Schioppa C. and J.A. Assad (2006), "Neurons in the orbitofrontal cortex encode economic value", *Nature*, 441: 223-226.
- Padoa-Schioppa C. and K.E. Conen (2017), "Orbitofrontal cortex : A neural circuit for economic decisions", *Neuron*, 96(4) : 736-754.
- Palmer J., A.C. Huk, and M.N. Shadlen (2005), "The effect of stimulus strength on the speed and accuracy of a perceptual decision", *Journal of Vision*, 5(5): 376-404.
- Parfit D. (1984), *Reasons and persons*, Oxford: Clarendon Press.
- Peirce C.S. (1878), "Deduction, induction, and hypothesis", *Popular Science Monthly*, 13: 470-482.
- Pelloux B., J-L. Rullière et F. van Winden (2009), « La neuroéconomie dans l'agenda de l'économie comportementale », *Revue Française d'Economie*, XXIII (avril) : 3-36.
- Pesendorf W. (2006), "Behavioral economics comes of age", Working Paper Princeton University.
- Pfeiffer U.J. (2013), "Oxytocin not always a moral molecule", *Frontiers in Human Neuroscience*, 7(10).
- Phelps E.A. (2009), "The study of emotions in neuroeconomics", in Glimcher P. *et al.* (eds.) (2009), chap.16: 233-250.
- Plassman H., J. O'Doherty, and A. Rangel (2007), "Orbitofrontal cortex encodes willingness to pay in everyday economic transactions", *Journal of Neuroscience*, 27: 9984-9988.
- Platt M.L. and P.W. Glimcher (1999), "Neural correlates of decision variables in parietal cortex", *Nature*, 400: 233-238.
- Platt M.L. and S.A. Huettel (2008), "Risky business: the neuroeconomics of decision making under uncertainty", *Nature Neuroscience*, 11: 398-303.
- Poldrack R.A. (2006), "Can cognitive processes be inferred from neuroimaging data?" *Trend in Cognitive Sciences*, 10(2): 59-63.

- Poldrack R.A., C.I. Baker, J. Durnez, K.J. Gorgolewski, P.M. Matthews, M.R. Munato, T.E. Nichols, J.B. Poline, E. Vul, and T. Yakoni (2017), "Scanning the horizon: Towards transparent and reproducible neuroimaging resource", *Nature Review Neuroscience*, 18(2): 115-126.
- Premack D. and G. Woodruff (1978), "Does the chimpanzee have a 'theory of mind'?" *Behavioral and Brain Sciences*, 4: 515-526.
- Preuschoff K., P. Bossaerts, and S.R. Quartz (2006), "Neural differentiation of expected reward and risk in human subcortical structures", *Neuron*, 51: 381-390.
- Price C.J. and K.J. Friston (2005), "Functional ontologies for cognition: the systematic definition of structure and function", *Cognitive Neuropsychology*, 22: 262-275.
- Quartz, S.R. (2008), "From cognitive science to cognitive neuroscience to neuroeconomics", *Economics & Philosophy*, 24: 459-471.
- Raichle M.E. (2015), "The brain's default mode network", *Annual Review of Neuroscience*, 38: 433-447.
- Rangel A. (2009), "The computation and comparison value in goal-directed choice", in Glimcher P. *et al.* (eds.) (2009), chap. 28: 425-440.
- Rangel A., C. Camerer, and P.R. Montague (2008), "A framework for studying the neurobiology of value-based decision making", *Nature Reviews*
- Rangel A and J.A. Clithero (2014), "The computation of stimulus values in simple choice", in Glimcher P. and E. Fehr (eds.) (2014), chap. 8: 125-148.
- Rangel A. and T. Hare (2010), "Neural computations associated with goal-directed choice", *Current Opinion in Neurobiology*, 20: 262-270.
- Ratcliff R., P.L. Smith, S.D. Brown, and G. McKoon (2016), "Diffusion decision model: current issues and history", *Trends in Cognitive Science*, 20: 260-281.
- Rawls J. (1987), *Théorie de la justice*, Paris: Editions du Seuil (trad. française de *A theory of justice*, The Belknap Press of Harvard University Press, 1971).
- Riedl R. and A. Javor (2012), "The biology of trust: integrating evidence from genetics, endocrinology and functional brain imaging", *Journal of Neuroscience, Psychology, and Economics*, 5(2): 63-91
- Rilling J.K., A.C. DeMarco, P.D. Hackett, R. Thompson, B. Ditzen, R. Patel, and G. Pagnoni (2012), "Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men", *Psychoneuroendocrinology*, 37: 447-461.
- Rilling J.K. and A.C. Sanfey (2011), "The neurosciences of social decision-making", *Annual Review of Psychology*, 62: 43-48.
- Rilling J.K., A.G. Sanfey, J.A. Aronson, L.E. Nystrom, and J.D. Doherty (2004), "The neural correlate of theory of mind within interpersonal interactions", *NeuroImage*, 22(4): 1694-1703.
- Rizzolatti G. and C. Sinigaglia (2011), *Les neurones miroirs*, Paris : Odile Jacob (trad. de l'ouvrage en italien publié chez Raffaello Cortina Editore, 2006).
- Roitman J.D. and M.N. Shadlen (2002), "Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task", *Journal of Neuroscience*, 22: 9475-9489.
- Rolls E.T. (2014), *Emotion and decision-making explained*, Oxford, UK: Oxford University Press.
- Rolls E.T. and Deco G. (2010), *The noisy brain: stochastic dynamics as a principle of brain function*, Oxford: Oxford University Press.
- Rose N. and J.M. Abi-Rached (2013), *Neuro: the new brain sciences and the management of the mind*, Harvard University Press.
- Ross D. (2005), *Economic theory and cognitive science: microexplanation*, Cambridge, MA: MIT Press.
- Ross D. (2008), "Two styles of neuroeconomics", *Economics and Philosophy*, 24(3): 473-483.
- Rubinstein A. (2006), "Comments on behavioral economics" in Blundell R., W.K. Newey, T. Persson (eds.), *Advances in economic theory (2005 World Congress of the Econometric Society)*, vol II, Cambridge University Press, 246-254.
- Rubinstein A. (2008), "Comments on neuroeconomics", *Economics & Philosophy*, 24: 485-494 (trad. française in *Revue Française d'Economie*, 2009, XXIII (avril) : 159-174).
- Ruff C.C. and E. Fehr (2014), "The neurobiology of rewards and values in social decision", *Nature Review Neuroscience*, 15: 549-562.
- Ruff C.C. and S.A. Huetzel (2014), "Experimental methods in cognitive neuroscience", in Glimcher P. et E. Fehr (eds.) (2014), chap. 6: 77-108.
- Ruff C.C., G. Ugazio and E. Fehr (2013), "Changing social norm compliance with noninvasive brain stimulation", *Science*, 342: 482-484.
- Sanfey A.G., G. Loewenstein, S.M. McClure, and J.D. Cohen (2006), "Neuroeconomics: cross-currents in research on decision-making", *Trends in Cognitive Sciences*, 10(3): 108-116.

- Sanfey A.G. and J.K. Rilling (2011), "Neural bases of social decision making", in Vartorian O. and D.R. Mandel (eds.) (2011).
- Sanfey A., J. Rilling, J. Aronso, L.E. Nystrom, and J.D. Cohen (2003), "The neural basis of economic decision-making in the ultimatum game", *Science*, 300: 1755-1758.
- Santos L.R. and M.L. Platt (2014), "Evolutionary anthropological insights into neuroeconomics: what non-human primates can tell us about human decision-making", in Glimcher P. and E. Fehr (eds.) (2014), chap. 7: 109-122.
- Schmidt C. et P. Livet (2014), *Comprendre nos interactions sociales. Une perspective neuroéconomique*, Paris : Odile Jacob.
- Schultz W. (2009), "Midbrain dopamine neurons: a retina of the reward system?" in Glimcher P. et al. (eds.) (2009), chap. 21: 323-329.
- Schultz W., P. Dayan, and P.R. Montague (1997), "A neural substrate of prediction and reward", *Science*, 275 (5306): 1593-1599.
- Schultz W., K. Preusschoff, C.F. Camerer, M. Hsu, C.D. Fiorillo, N.P. Tobler, and P. Bossaerts (2008), "Explicit neural signals reflecting reward uncertainty", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363: 3801-3811.
- Serra D. (2012a), « Un aperçu historique de l'économie expérimentale : des origines aux évolutions récentes », *Revue d'Economie Politique*, 122 (5) : 749-786.
- Serra D. (2012b), « Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble », *Revue de Philosophie Economique*, 13 (1) : 21-78.
- Serra D. (2015), *Histoire de l'économie expérimentale*, Saarbrücken : EUE.
- Serra D. (2016), *Neuroéconomie*, Paris : Economica.
- Serra D. (2017), *Economie comportementale*, Paris : Economica.
- Shamay-Tsoory S.G., A. Abu-Akel, S. Palgi, et al. (2013), "Giving peace a chance: oxytocin increases empathy to pain in the context of the Israeli-Palestinian conflict", *Psychoneuroendocrinology*, 38(12): 3139-44.
- Siegel J.Z. and M.J. Crockett (2013), "How serotonin shapes moral behavior", *Annals of the New York Academy of Sciences*, 1299(1): 42-51.
- Singer T. and A. Tusche (2014), "Understanding others: brain mechanisms of theory of mind and empathy", in Glimcher P. and E. Fehr (eds.) (2014), chap. 27: 513-532.
- Sinnott-Armstrong W (1996), "Moral scepticism and justification", in Sinnott-Armstrong W. and M. Timmons (eds.) (1996), *Moral Knowledge? New readings in moral epistemology*, Oxford: Oxford University Press.
- Sinnott-Armstrong W., A. Roskies, T. Brown, and E. Murphy (2008), "Brain images as legal evidence", *Episteme: A Journal of Social Philosophy* (special issue: Evidence and Law), 5(3): 359-373.
- Smith V.L. (2003), "Constructivist and ecological rationality in economics", in Frängsmyr T. (ed.), *Les prix Nobel Prizes 2002*, Nobel Foundation, Stockholm (reproduit in *American Economic Review*, 93: 465-508)
- Smith V.L. (2008), *Rationality in economics: constructivist and ecological forms*, Cambridge: Cambridge University Press.
- Smith A., B.D. Bernheim, C. Camerer, and A. Rangel (2014), "Neural activity reveals preferences without choices", *American Economics Review: Microeconomics*, 6(2): 1-36.
- Stanton A.A. (2009), « Commentaires sur "neuroéconomie : une relecture critique" », *Revue Française d'Economie*, XXIII (avril) : 139-157.
- Stoop R. (2014), "Neuromodulation by oxytocin and vasopressin in the central nervous system as a basis for their rapid behavioural effects", *Current Opinion in Neurobiology*, 29: 187-193.
- Taya F. (2012), "Seeking ambiguity: a review on neuroimaging studies on decision making under ambiguity", *Recherches économiques de Louvain*, 78(3): 83-100.
- Teller D.Y. and E.N. Pugh Jr. (1983), "Linking propositions in color vision", in Mollon J.D. & T. Sharpe (eds.), *Colour vision: Physiology and Psychophysics*, New York: Academic Press, 11-21.
- Thaler R.H. (1985), "Mental accounting and consumer choice", *Marketing Science*, 4: 199-214.
- Thaler R.H. (1999), "Mental accounting matters", *Journal of Behavioral Decision Making*, 12: 183-206.
- Tirole J. (2007), Interview, *La Recherche*, n° 414 (décembre).
- Tverski A. and D. Kahneman (1992), "Advances in prospect theory: cumulative representation of uncertainty", *Journal of Risk and Uncertainty*, 5: 297-323.
- Uttal W. (2001), *The new phrenology: the limits of localizing cognitive processes in the brain*, Cambridge: MIT Press.

- Vallois N. (2011), "The pathological paradigm of neuroeconomics", *Oeconomia*, 1(4): 525-556.
- Vallois N. (2012), *L'émergence d'un nouveau domaine du savoir : la neuroéconomie*, Thèse de doctorat de sciences économiques, Université Paris 1 Panthéon-Sorbonne (dir. : A. Cot), 455 p.
- Van Rooij M. and G. Van Orden (2011), "It's about space, it's about time, neuroeconomics and the brain sublime", *Journal of Economic Perspectives*, 25(4): 31-56.
- Vercoe M. and P.J. Zak (2010), "Inductive modelling using causal studies in neuroeconomics: brains and drugs", *Journal of Economic Methodology*, 17(2): 133-146.
- Vromen J. (2007), "Neuroeconomics as a natural extension of bioeconomics: the shifting scope as standard economic theory", *Journal of Bioeconomics*, 9(2): 145-167.
- Vromen J. (2011), "Neuroeconomics: two camps gradually converging: what can economics gain from it?" *International Review of Economics*, 58: 267-285.
- Wager T.D., M.A. Lindquist, and L.A. Kaplan (2007), "Meta-analysis of functional neuroeconomics data: current and future directions", *Social Cognitive and Affective Neuroscience*, 2(2): 150-158.
- Wallis J.D. (2012), "Cross-species studies of orbitofrontal cortex and value-based decision-making", *Nature Neuroscience*, 15(1): 13-19
- Walliser B (2012), "From psycho-economics to neuro-economics", *Recherches économiques de Louvain*, 78(3) : 13-28.
- Wibral M., T. Dohnen, D. Kingmüller, B. Weber, and A. Falk (2012), "Testosterone administration reduces lying in men", *PLoS ONE*, 7(10): e46774.
- Wilkinson N. (2008), *An introduction to behavioural economics*, Palgrave Macmillan.
- Wilson E.O. (1975), *Sociobiology, the new synthesis*, Harvard University Press.
- Wong S.W., G. Xue, and A. Bechara (2011), "Integrating fMRI with psychophysiological measurements in the study of decision making", *Journal of Neurosciences, Psychology, and Economics*, 4: 85-94.
- Yarkoni T., R.A. Poldrack, T.E. Nichols, D.C. Van Essen, and T.D. Wagner (2011), "Large-scale automated synthesis of human functional neuroimaging data", *Nature Methods*, 8(8): 665-670.
- Zak P.J. (2004), "Neuroeconomics", *Philosophical Transactions of the Royal Society B: Biological Science*, 359(1451): 1737-1748.
- Zak P.J. (2011), "The physiology of moral sentiments", *Journal of Economic Behavior and Organization*, 77: 53-65.
- Zak P.J., K. Borja, W.T. Matzner, and R. Kurzban (2005a), "The neuroeconomics of distrust: sex differences in behavior and physiology", *American Economic Review*, 2005(2): 360-363.
- Zak P.J., R. Kurzban, and W.T. Matzner (2004), "The neurobiology of trust", *Annals of the New York Academy of Sciences*, 1032: 224-227.
- Zak P.J., R. Kurzban, and W.T. Matzner (2005b), "Oxytocin is associated with human trustworthiness", *Hormones and Behavior*, 48: 522-527.
- Zak P.J., R. Kurzban, S. Ahmadi, R.S. Swerdloff, J. Park, *et al.* (2009), "Testosterone administration decreases generosity in the ultimatum game", *PLoS ONE*, 4: e8330.
- Zak P.J., A. Stanton, and S. Amadi (2007), "Oxytocin increases generosity in humans", *PLoS ONE*, 2: e1128.

CEE-M Working Papers¹ - 2019

- WP 2019 - 01 **Adrien Nguyen-Huu. & Antonin Pottier**
«Hicksian Traverse Revisited: Conditions for the Energy Transition»
- WP 2019 - 02 **Oumar Mbodji, Adrien Nguyen-Huu & Traian A. Pirvu**
«Optimal Sharing Rule for a Household with a Portfolio Management Problem»
- WP 2019 - 03 **Tristan Cotty, Tristan, Elodie Maitre d'Hotel & Julie Subervie**
«Inventory credit to enhance food security in Africa»
- WP 2019 - 04 **Raphael Soubeyran**
«Incentives, Pro-social Preferences and Discrimination»
- WP 2019 - 05 **Adrien Fabre, Mouez Fodhaz & Francesco Ricci**
«Optimal Timing of Energy Production»
- WP 2019 - 06 **Claude Bismut & Ismaël Rmajo**
«A world of low interest rates»
- WP 2019 - 07 **Gabriela Demarchi, Julie Subervie, Fernando Palha Leite & Jean-Paul Laclau**
« Farmers' preferences for water-saving strategies in brazilian eucalypt plantation »
- WP 2019 - 08 **Arnaud Tognetti, David Doat, Dimitri Dubois & Rustam Romaniuc**
« Does the presence of a physically disabled erson in the group increase cooperation? An experimental test of the empaty altruism hypothesis »
- WP 2019 - 09 **Isabelle Tritsch, Gwenole Le Velly, Benoit Mertens, Patrick, Meyfroidt, Christophe Sannier, Jean-Sylvestre Makak & Kenneth Hounbedji**
« Do Forest-Management Plans and FSC Certification Reduce deforestation in the Congo Basin? »
- WP 2019 - 10 **Jean-Michel Salles**
« Valuing the loss and damage from climate change: a review of some current issues »
- WP 2019 - 11 **Can Askan Mavi**
« Can harmful events be another source of environmental traps? »
- WP 2019 - 12 **Daniel Serra**
« Neuroeconomics and Modern Neuroscience»

¹ CEE-M Working Papers

Contact : laurent.garnier@inra.fr/ 04 99 61 31 21

WP 2019 - 13

Daniel Serra

« La neuroéconomie en question : débats et controversies »