

Neurosciences appliquées : une question de limites

Jean-Luc PETIT

Université de Strasbourg, Faculté de philosophie

CREPhAC

Tandis que la physique appliquée ou la chimie appliquée désignent des innovations qui changent notre vie, on peut se demander si les neurosciences sociales, la neuroéconomie ou la neuroéthique font plus qu'un recodage en termes neuroscientifiques des doctrines connues de disciplines existantes. Sous l'apparence académique d'une dispute de frontières entre les disciplines se profile ici le sérieux du conflit des ontologies : le monisme naturaliste des neurosciences ne va-t-il pas bientôt suppléer le pluralisme épistémologique des sciences sociales avec en guise de société ou d'économie, les événements corrélatifs dans les circuits baso-amygdalo-frontaux du cerveau humain ?

Je commencerai par me référer au paradigme galiléen, en faisant appel à GALILÉE en tant qu'inventeur de l'idée d'une science de la nature rigoureusement *exacte*. Cela signifie que les formes typiques à contenus sensoriellement qualifiés, les mouvements perçus comme efforts limités vers des buts, qui composent notre expérience quotidienne, soient tenus pour autre chose que ce qu'ils sont pour nous; à savoir pour les indices des formes pures de la géométrie et des mouvements indéfiniment conservés de la mécanique. Sous l'apparence édifiante d'une simple observation des faits, seule retenue par la pédagogie, ce collapse du monde vécu en domaine de la mesure résulte d'une opération absolument inédite dont GALILÉE a assumé la responsabilité historique : une objectivation combinant idéalisation, abstraction, formalisation, construction, substruction, bien décrites par Husserl dans *La Crisis*¹. C'est l'origine de la Mécanique rationnelle, domaine propre d'objectivités théoriques s'étendant à l'infini, et néanmoins fermé sous un système d'axiomes pour un traitement formel a priori par les opérations numériques et géométriques (*cum fundamento in re*). On peut prendre comme exemple la loi de la chute des corps, d'après l'histoire de sa découverte restituée en particulier par Alexandre KOYRE² dans la ligne de HUSSERL.

¹ E. Husserl, *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie: Eine Einleitung in die phänomenologische Philosophie*, II. §. 9. Galileis Mathematisierung der Natur.

² Koyré A., La loi de la chute des corps, *Revue philosophique*, 1937, 152-172.

En contradiction avec l'intuition et avec ARISTOTE, une feuille d'arbre et une bille de plomb en chute libre *doivent nécessairement* toucher le sol en même temps. Pourquoi ? Parce que le mouvement ne dépend pas de la nature particulière des mobiles, que l'espace géométrique et le temps d'horloge en sont les seuls paramètres pertinents et que leur loi de composition la plus intelligible est que le mouvement des deux corps subit une même accélération uniforme et proportionnelle au temps de chute. Au temps de chute et non pas à la distance parcourue à partir du point de départ, ce que GALILÉE affirme néanmoins par préjugé de géomètre, faisant preuve d'un géométrisme intempérant qui lui a fait manquer l'énoncé correct de la loi : « Je suppose et peut-être pourrais-je démontrer que le grave tombant naturellement va en augmentant continuellement sa vitesse dans la mesure où s'accroît la distance du point dont il est parti.³ »

Sur cette toile de fond épistémologique, je me réfère maintenant au modèle mathématique de la décision rationnelle qui va nous occuper dans cet exposé. En référence à la naissance de la physique, VON NEUMANN et MORGENSTERN⁴ fondent une science des interactions humaines qu'ils veulent rigoureusement *exacte*. A cette fin la relation complexe entre le comportement des individus et le marché est soumise à une opération de mathématisation essentiellement de même nature que celle de GALILÉE, selon l'opinion des auteurs eux-mêmes. Il s'agit, en effet, d'une opération d'idéalisation, d'abstraction, de formalisation, de construction, de substruction, etc.. On simplifie en particulier la psychologie de la motivation des agents économiques en assumant avec l'utilitarisme classique que la valeur est convertible en monnaie. Les quantités de biens de consommation ou d'échange sont censées être commensurables en termes d'unités toujours identiques à elles-mêmes, comparables entre elles, divisibles, transférables et représentables par un nombre qu'on appelle le maximum d'utilité. On dégage de même les propriétés formelles de l'interaction sociale en posant l'identité, et non pas seulement l'analogie entre les systèmes sociaux et des jeux de stratégie (et non les jeux de hasard). Car ce qui complique l'interaction sociale par rapport à une économie de Robinson Crusoë, lequel n'a affaire qu'aux alea de l'environnement naturel, c'est que l'action de l'un dépend de ce que l'autre va faire en retour. La situation devient plus claire si l'on conçoit cette interaction comme un jeu de stratégie, puisque la stratégie choisie par un joueur dépend très généralement de son anticipation de la stratégie du partenaire. Certaines

³ Galilée, *Frammenti attenenti ai Discorsi, Opera* vol.VIII, p. 373.

⁴ Von Neumann J. & Morgenstern O., *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1944.

caractéristiques formelles de l'interaction ressortent de cette nouvelle façon de voir les choses. Par exemple, l'existence dans certains jeux (pas dans tous !) d'une stratégie dominante : la stratégie la plus avantageuse, quelle que soit la décision du partenaire, est une propriété formelle – démontrable a priori – du système d'interactions que constitue le jeu, qui ne doit rien à la psychologie du comportement du joueur individuel.

Si je réfléchis maintenant à la critique – largement vulgarisée aujourd'hui – du modèle d'agent rationnel et aux ambiguïtés de cette critique, il y a bien sûr une apparence de légitimité de l'impératif de réalisme pour une science du comportement économique qui prétend apporter une contribution utile à la solution des questions brûlantes de la vie quotidienne : crises, inégalités, chômage, pauvreté, etc. Cependant, il y a également un risque de mécompréhension du statut épistémologique de l'usage de modèles en science. Par définition, un modèle ne peut pas coller à la réalité, son intérêt tenant au fait qu'il décolle du réel précisément pour en dégager les seuls aspects ou propriétés qu'on peut (ou qu'on veut) expliquer à l'exclusion de tous les autres. D'où le mirage qu'est souvent la promesse d'un retour de l'abstraction du modèle de calculateur rationnel au concret de la décision de l'individu. Non, sans doute, la décision de tel ou tel individu dans son contexte social particulier, laquelle est laissée à sa contingence. Mais une décision non moins théorique, en fait, que celle du modèle classique, sauf qu'elle serait fondée sur les seules ressources des mécanismes fonctionnels connus du cerveau, où la contribution des émotions tend à éclipser celle du raisonnement logique. C'est ainsi qu'on en vient au programme des neurosciences : les émotions guident la décision. Je souhaiterais ici faire partager mon sentiment qu'il existe un contraste majeur entre une science économique adulte, qui touche les limites de son paradigme de rationalité et tente de repartir de la psychologie des décideurs empiriques, si nuls en matière de finances qu'ils soient en moyenne⁵, et, d'autre part, des neurosciences trop jeunes encore pour ne pas croire que le monde, le vaste monde, leur est ouvert.

Je discuterai la proposition d'Antonio DAMASIO : corps et émotions recèlent-ils vraiment des principes suffisants pour sauvegarder l'adéquation de l'action à l'intention de l'agent comme à la situation environnante, ce qui est typique d'un comportement rationnel ? Dans *L'erreur de Descartes*, DAMASIO soutient que les émotions guident la décision parce qu'elles signalent les conséquences prévisibles de l'action par leurs effets bien connus sur le

⁵ Frydman C. & Camerer C.F., The Psychology and Neuroscience of Financial Decision Making, *Trends in Cognitive Sciences*, 1592, 2016, 1-15.

corps⁶. Saurait-on, sans elles, comme cela fait mal de perdre de l'argent ? C'est une théorie fondée sur le syndrome du lobe frontal des patients qui sont porteurs d'une lésion du cortex préfrontal ventromédian (il inclut aussi les lésions des noyaux amygdaliens), et qui manifestent au niveau comportemental une défaillance du jugement et de la prise de décision dans la vie quotidienne, tandis qu'ils éprouvent des émotions et des sentiments qui sont régulièrement en désaccord avec leurs situations. Au test du jeu de cartes ces patients manifestent une préférence anormale pour les paquets de cartes qui rapportent un gain supérieur immédiat, mais aussi une perte supérieure à long terme. Au demeurant, on se souvient que cette proposition a bénéficié d'une opération de promotion médiatique fort réussie : les DAMASIO publiant simultanément dans la revue *Science*, sous le titre *Le retour de Phineas GAGE, le crâne d'un patient célèbre donne la clé du cerveau*, une interprétation par le syndrome du lobe frontal d'un cas des annales médicales⁷. Phineas GAGE, employé du chemin de fer aux États-Unis, a été accidenté en 1848 : une barre à mine lui a traversé le crâne. Sur la reconstitution par ordinateur des aires anatomiques de son cerveau à partir de l'os crânien perforé, il semble que les lésions concernent principalement le cortex orbitofrontal et non les aires fonctionnelles du langage ni les aires motrices ou somatosensorielles. Ce qui explique l'exceptionnelle survie du malheureux – une douzaine d'années – malgré des écarts de conduite que DAMASIO croit pouvoir attribuer à la perte d'influence régulatrice des émotions sur la prise de décision.

Pour l'historien de la philosophie, bien sûr, DESCARTES n'est nullement en cause en ce débat : il suffit de lire le *Traité des passions* pour se persuader qu'il n'était pas ce rationaliste abstrait dont on a fait un repoussoir tellement commode. Il s'agit plutôt, en l'occurrence, pour le neurophysiologiste, de fonder sur la connaissance des mécanismes cérébraux la critique du modèle normatif de la décision rationnelle, comme calcul du maximum d'utilité sans considération pour les risques encourus ; une critique déjà formulée, mais en termes purement psychologiques, par KAHNEMAN et TVERSKY⁸. Selon DAMASIO, la différence des réactions somatiques face à des événements certains et à des événements incertains pourrait expliquer la fréquence dans les comportements des sujets en situation de choix économique d'une alternance irrationnelle entre la prise de risques et l'aversion aux risques.

⁶ Damasio A.R., *Descartes' Error: Emotions, Reason and the Human Brain*, Grosset Putnam, New York.

⁷ Damasio H., Grabowski T., Frank R., Galaburda A.M., Damasio A.R., The return of Phineas Gage: clues about the brain from the skull of a famous patient, *Science*, 264, 5162, 1994, 1102-1105.

⁸⁸ Tversky, A., Kahneman, D., The framing of decisions and the rationality of choice, *Science* 211, 1981, 453–458.

D'après la théorie des marqueurs somatiques, lorsque la bouilloire des émotions est relativement peu agitée, le *feedback* somatique est modéré et peut exercer une influence avantageuse sur la décision à prendre, grâce à un effet régulateur sur l'activité des systèmes cérébraux des émotions et de l'action : l'amygdale, les noyaux du tronc cérébral, le cortex ventromédian et l'insula. Mais en cas de crash boursier, les porteurs d'actions se précipitent pour vendre ; tandis que s'ils sont surpris par une performance imprévue ils se précipitent pour acheter. Ces comportements irrationnels s'expliqueraient par le fait que la bouilloire des émotions entrant en ébullition, le *feedback* somatique est exagéré et ses effets sur les systèmes de la décision sont potentiellement catastrophiques⁹.

Disons tout de suite que le fait – à supposer que ce soit un fait – que des émotions modérées favorisent des décisions avantageuses, tandis que la perte de contrôle émotionnel peut avoir des conséquences dommageables, ce truisme ne prouve pas que le principe régulateur de la conduite soit lui-même de nature émotionnelle. Il demeure au moins concevable que le jugement, et plus précisément le calcul de l'utilité, soit l'instance responsable à la fois de la décision et du tempérament des émotions qui permet qu'elle soit avantageuse. L'ignorer, c'est retomber sur la doctrine de Hume : l'apparente impuissance de la raison une fois qu'on a mis au compte des passions toute la dimension énergétique de la vie psychique. Incapables, par elles-mêmes, de trancher le différent entre ce monisme émotionnel et le dualisme cognitif, les neurosciences pourront-elles jouer le rôle d'arbitre que DAMASIO leur assigne sur la question du réalisme du modèle de décideur rationnel ?

Une série de travaux en imagerie fonctionnelle cérébrale se sont appliqués depuis une dizaine d'années à la validation de son hypothèse, sans toutefois aboutir à une conclusion incontestée sur l'essentialité de la participation des émotions à la décision. Une grande partie d'entre eux ne concerne d'ailleurs pas directement la critique du modèle d'agent rationnel puisqu'ils placent les sujets face à une décision en situation ambiguë dans un contexte de jeu de hasard et non de jeu de stratégie. Il est notoire que la plupart du temps on prend une décision sans connaître la mesure de probabilité exacte du résultat du choix. Dans une recherche conduite par le groupe de Colin CAMERER à l'Institut de Technologie de Californie¹⁰, les sujets ont eu le choix entre un gain certain et un pari de deux types : un pari avec une information suffisante et un pari sans information suffisante pour avoir une mesure de probabilité sur l'issue du choix.

⁹ Bechara A. & Damasio A.R., The somatic marker hypothesis: A neural theory of economic decision, *Games and Economic Behavior* 52, 2005, Fig. 10, 366.

¹⁰ Hsu M., Bhatt M., Adolphs R., Tranel D., Camerer C.F., Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making, *Science*, vol. 310, 2005, 1680-1683.

Les régions du cortex orbitofrontal et des amygdales sont plus activées par la décision en condition d'ambiguïté, c'est-à-dire de méconnaissance de la mesure de probabilité, qu'en condition d'information suffisante sur la probabilité du résultat. (Ce résultat est toutefois controversé). En revanche, le noyau caudé qui fait partie du striatum dorsal semble également plus actif en condition d'information suffisante.

L'hypothèse validée par cette expérimentation est qu'il existerait un système biologique d'alerte émotionnelle précoce lorsque l'environnement ne fournit pas une information suffisante pour le choix et que la décision peut avoir des conséquences dangereuses. A peine validée cette hypothèse a été contestée, notamment par l'équipe de RUSTICHINI et al., de l'Université du Minnesota à Minneapolis¹¹, qui a refait une manipulation sur la décision – entendue comme acte cognitif cette fois – remarquant que le test du jeu de cartes de DAMASIO mélangeait le processus mental de la décision, lui-même, avec l'émotion qu'on peut éprouver du fait d'un gain ou d'une perte. Si on n'informait pas les sujets en cours d'expérimentation du résultat de leur choix, les auteurs observent qu'on aurait plus de chance de tester le processus de décision en lui-même, indépendamment de l'émotion. Sur cette hypothèse, ils ont proposé une nouvelle tâche consistant pour les sujets à choisir une boule dans une urne contenant des boules rouges ou des boules bleues en proportions variées, et qui rapportent des gains différents, dans des conditions où l'information donnée permet ou ne permet pas une estimation précise des loteries qu'il faut comparer.

Les régions associées à l'influence des émotions sur la décision : le cortex orbitofrontal et l'amygdale, dont on vient de parler, ne sont pas activés dans ces conditions par le choix entre loteries en condition d'incertitude, tandis que les régions activées sont le cortex préfrontal, qu'on associe à la planification de l'action, et le lobe pariétal, plus précisément, le précuneus, qu'on associe à une estimation des quantités visuospatiales et non verbales, quantités au sens quasi numérique. L'hypothèse étant, cette fois-ci une hypothèse cognitiviste et non pas émotionnaliste, que l'activité mentale recrutée par la décision en condition d'incertitude et sans mesure possible de probabilité du résultat, n'est pas de nature émotionnelle, mais serait plutôt une activité qui mérite le titre de cognitive à l'égal du jugement explicite. C'est-à-dire un calcul mental approximatif.

¹¹ Rustichini A., Dickhaut J., Ghirardato P., Smith K., Pardo J.V., A brain imaging study of the choice procedure, *Games and Economic Behavior*, 52, 2005, 257-282.

Passons de la question de la suffisance causale, comme ci-dessus, à celle de la nécessité causale, pour la décision économique, du mécanisme des marqueurs somatiques et donc du fonctionnement normal du cortex frontal médian et des amygdales, leurs substrats cérébraux. Voici deux contre-exemples dont on pourra toujours dire qu'ils ont un caractère anecdotique, dans la mesure où ce sont des cas singuliers et non des régularités statistiquement fondées sur des cohortes entières de patients. Il s'agit d'une part d'un cas de lésion du cortex frontal médian avec aptitude normale à « la mentalisation ». La mentalisation est le jargon des psychologues pour dire que non seulement nous faisons des inférences sur les états mentaux d'autrui toutes les fois que nous entrons en interaction avec autrui, mais que ces inférences rendent possible cette interaction – une croyance commune en psychologie. Un sujet (G.T.) qui a une lésion étendue dans la région frontale médiane consécutive à un infarctus de l'artère cérébrale antérieure, pouvait néanmoins répondre de manière correcte aux tests de *Theory of Mind*, qui prouvent que l'on est capable de comprendre le comportement des protagonistes du récit d'une interaction sociale fictive sur la base des états mentaux qu'on leur attribue¹². Deuxième contre-exemple, je signalerai d'autre part un cas de perception normale de la peur dans les expressions faciales malgré une lésion bilatérale de l'amygdale. La personne (S.M.) avait ses deux amygdales sclérosées et cependant, pourvu qu'on guidât son regard vers les yeux sur l'image d'expression faciale proposée, elle se révélait capable de reconnaître la peur alors qu'elle semblait ne pas le pouvoir en vision naturelle¹³. Ce qui suggère que la difficulté éprouvée par cette patiente concernerait plutôt l'orientation de son attention visuelle et n'aurait pas de lien essentiel avec la reconnaissance des émotions. Donc l'amygdale n'est peut-être pas nécessairement la base du marquage émotionnel des stimuli et de l'information cognitive. C'est une question en tout cas.

J'aborde ensuite l'interaction sociale, telle que le modèle classique la conceptualise comme jeu de stratégie. L'émotion au sens de DAMASIO donne une interaction essentiellement symétrique : contagion, émulation, empathie, etc. On fait la même chose. Identification passive, peut-être, mais pas réciprocation active. Or, normalement l'interaction sociale est asymétrique : c'est un déséquilibre dynamiquement compensé. Dans l'interaction, chaque participant sait a priori quelle action, différente de l'action observée, il doit accomplir

¹² Bird C.M., Castelli F., Malik O., Frith U., Husain M., The impact of extensive medial frontal lobe on 'Theory of Mind' and cognition, *Brain*, 127, 2004, 914-928..

¹³ Adolphs R., Gosselin F., Buchanan T.W. Tranel D., Schyns P., Damasio A.R., A mechanism for impaired fear recognition after amygdala damage, *Nature*, 433, 2005, 68-72.

en réponse à l'action observée, afin de satisfaire l'exigence normative créée par l'initiateur de l'acte social¹⁴ :

Donner – recevoir ; demander – accorder ; commander – obéir ; promettre – faire valoir une prétention sur la chose promise ; se repentir – accorder son pardon ; requérir – se présenter devant les autorités ; investir – verser les dividendes, et ainsi de suite.

De cette asymétrie essentielle de l'action sociale, les jeux de stratégie sont, comme l'avaient bien vu les auteurs de l'ouvrage de référence : *Theory of games and economic behavior*, un meilleur modèle que la contagion des neurosciences affectives. Le sentiment que la recherche des fondements cérébraux de l'interaction sociale ne devait pas être limitée aux symétries de l'interaction émotionnelle a pu entrer dans la détermination du programme de la neuroéconomie, l'imagerie cérébrale des protagonistes d'un jeu de stratégie modèle des relations sociales fiduciaires du marché. Faute de pouvoir passer en revue une littérature d'ores et déjà considérable, je me limiterai à présenter succinctement trois recherches qui ne sont pas d'ailleurs des plus récentes, car plus on avance, plus l'affinement du protocole expérimental complique le compte-rendu des résultats et plus le philosophe est à la peine dans son effort de compréhension et d'interprétation.

Je ferai état tout d'abord du travail le plus ancien, réalisé par une équipe de l'Université de l'Arizona à Tucson¹⁵, où l'on place l'un des deux sujets, le décideur D1, dans le scanner d'imagerie fonctionnelle cérébrale tandis que le partenaire, le décideur D2, reste à l'extérieur, dans un jeu consistant à se répartir des gains monétaires à tour de rôle. La confiance en la réciprocité du choix futur du partenaire de jeu est supposée motiver la préférence du sujet pour un choix coopératif. Cette aptitude est censée reposer sur une évaluation de l'état d'esprit du partenaire et sa mise en relation avec une expectative de gains futurs supérieurs aux gains immédiats. On distingue deux catégories de sujets sur la base des comportements observés : des sujets dits « coopérants », qui manifestent une préférence naturelle pour la coopération, et des sujets « non coopérants », qui font comme si le point de vue de l'autre était sans importance. Les sujets coopérants mobilisent un réseau qui comporte : le cortex préfrontal médian, associé classiquement d'une part à la « théorie de l'esprit », d'autre part à l'attention reportée sur les gains futurs, doublée de l'inhibition du désir de gains immédiats. Mais aussi le cortex pariétal,

¹⁴ L'essentielle asymétrie des actes sociaux fonde la possibilité d'une théorie a priori du droit (Reinach A., *Les fondements a priori du Droit Civil*, trad. R. de Calan, Vrin, 2004).

¹⁵ McCabe K., Houser D., Ryan L., Smith V., Trouard T., A functional imaging study of cooperation in two-person reciprocal exchange, *PNAS*, 98/20, 2001, 11832-11835.

voie visuelle du « où ? » : « de quel côté dois-je tendre la main » ? Sans oublier le thalamus, qui est le cerveau sensoriel. A l'opposé, les activités cérébrales des sujets non coopérants n'accusent pas de différence significative entre le jeu contre un humain et le jeu contre un ordinateur. L'explication est que ces sujets appliquent une règle uniforme sans chercher à comprendre le caractère du partenaire. Le contraste entre les deux catégories de décideurs tend à confirmer que l'option de coopérer avec un partenaire dans l'espoir d'un mutuel bénéfice plutôt que d'adopter une tactique égoïste active un réseau neural intégrant ensemble la capacité de reconnaissance d'autrui avec le contrôle du désir de récompense hédonique.

Je passe au second travail, réalisé par James RILLING de l'Université de Princeton avec une équipe de l'Université Emory à Atlanta sur le dilemme du prisonnier réitéré¹⁶. De chacun des couples de joueurs un seul est passé au scanner d'imagerie fonctionnelle cérébrale. Chaque joueur a le choix de coopérer ou de ne pas coopérer, les gains en dollars dépendant de l'interaction entre les joueurs. Les joueurs choisissent leur stratégie à l'insu l'un de l'autre, mais la matrice des choix respectifs leur est révélée avant une prise d'image. Les comportements réciproques induisent l'activation du réseau de ce qu'on appelle la récompense dopaminergique parce que le neuromédiateur dopamine serait associé à la « récompense » – quoique sans doute pas au plaisir, mais plutôt à l'incitation à l'action et à la prolongation de l'action une fois initiée. Ce réseau dopaminergique comporte le striatum ventral antérieur, le cortex cingulaire antérieur, le cortex orbitofrontal et le cortex ventromédian. L'hypothèse est qu'il y a un effet gratifiant – qui détermine l'orientation des décisions ultérieures – de l'anticipation de la coopération future d'autrui combiné avec une inhibition de la tendance égoïste à profiter de la générosité d'autrui sans compensation. La décision de coopérer après un choix coopératif du partenaire active un réseau qui inclut les régions activées suite à une partie avec coopération mutuelle. On retrouve ici un résultat classique, l'action symbolique (la disposition à persister dans un choix coopératif) emprunte les mêmes circuits cérébraux que l'action effective. Les auteurs concluent à l'existence au fondement de la division sociale du travail d'un processus de marquage hédonique des conséquences positives futures d'un choix actuel sur les choix ultérieurs du partenaire dans la poursuite d'un jeu de stratégie étendu sur plusieurs parties. Une interprétation, notons-le, qui irait dans le sens de la théorie des marqueurs somatiques de DAMASIO. Mais alors : marqueurs somatiques, ou calcul d'utilité ? En fait, la question est indécidable. Parce que le patron d'activations observé dans le cortex ventromédian, zone de

¹⁶ Rilling J.K., Gutman D.A., Zeh T.R., Pagnoni G., Berns G.S., Kilts C.D., A Neural basis for social cooperation, *Neuron*, 35, 2002, 395-405.

convergence des influences cortico-sous-corticales, peut toujours s'interpréter en termes de calcul. Pour maximiser ses gains sur plusieurs parties, il est en effet préférable d'encourager le partenaire à la coopération en se montrant coopérant, plutôt que le provoquer à la défection en se montrant égoïste.

Comme dernier exemple, j'évoquerai une situation de jeu de stratégie plus typiquement économique, dans une expérimentation en imagerie fonctionnelle cérébrale qui, cette fois, prend en considération les cerveaux des deux joueurs et qui recherche d'éventuelles corrélations entre les activités des deux cerveaux induites par les conditions du jeu. Cette recherche a été conduite par l'équipe de développement de l'Hyperscan au Baylor College of Medicine à Houston avec la collaboration de Colin CAMERER, de l'Institut de Technologie de Californie¹⁷. On peut intituler ce travail genèse de l'intention, l'intention en particulier de réinvestir, et formation de la réputation, plus précisément la bonne réputation de l'investisseur généreux. Chaque sujet a une tâche spécifique : il y a l'investisseur et l'entrepreneur. L'investisseur verse à l'entrepreneur une somme d'argent qui rapporte ; l'entrepreneur le rétribue en versant des dividendes suivant sa confiance en la générosité des réinvestissements futurs qu'il espère de l'investisseur. On le voit, les chercheurs ont élaboré des tests assez fins du comportement économique, même si l'approche est encore très restrictive et très réductrice. Est caractérisé comme bienveillant l'investisseur qui va augmenter son investissement en dépit de la mesquinerie manifestée par l'entrepreneur quand celui-ci ne rend pas suffisamment de dividendes. On considère comme malveillant de la part de l'investisseur de répondre à la générosité de l'entrepreneur par de la lésinerie en matière de réinvestissements. Quand on examine en imagerie fonctionnelle le cerveau de l'entrepreneur, il s'avère que c'est le noyau caudé qui réagit à la générosité de l'investisseur par un signal, base présumée de l'intention de faire confiance à l'investisseur à la prochaine partie.

On parle d'hyperscan quand les deux sujets passent au scanner et que l'on peut établir des corrélations entre les changements d'activités cérébrales de l'un et les changements d'activités cérébrales de l'autre et localiser les régions où les activités des deux cerveaux entrent en corrélation mutuelle. Les activations respectives des deux cerveaux montrent une corrélation entre le noyau caudé, le cortex cingulaire antérieur chez l'entrepreneur et d'autre part le cortex cingulaire moyen chez l'investisseur. Comment l'interpréter ? Le noyau caudé faisant partie du

¹⁷ King-Casas B., Tomlin D., Anen C., Camerer C.F., Quartz S.R., Montague P.R., Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange, *Science* 308, 2005, 78-83.

striatum dorsal, débouché du système dopaminergique de la récompense, on peut considérer que c'est un signe positif, un signe de confiance qui se manifeste par-là. Au cours du jeu, le pic de la corrélation intercérébrale se déplace dans le temps. Le signal, qui était d'abord, dans les parties initiales, un signal tardif de réaction à la connaissance de la décision de l'investisseur, devient au cours du jeu un signal précoce d'anticipation ou d'intention exprimant la confiance de l'entrepreneur dans le comportement de l'investisseur. L'interprétation est que le cerveau de l'entrepreneur développe au fur et à mesure que le jeu se déroule ce qu'on appelle un « modèle interne » de la fiabilité de l'investisseur, qui est probablement la base neurale de l'intention de la part de l'entrepreneur de faire confiance dans le futur à l'investisseur. Le signal cérébral ainsi identifié constitue donc un test assez fin de l'état d'esprit, de la disposition mentale sous-tendant la décision économique.

Je vais maintenant essayer d'interpréter ces données en termes plus philosophiques. Jusqu'à présent, la neuroéconomie s'est limitée, autant que je sache, à l'imagerie fonctionnelle du cerveau de l'individu – ou des deux individus – dans une tâche de jeu à deux. En guise de fondements cérébraux de la décision économique, on a repéré certaines traces cérébrales des attitudes dispositionnelles de l'individu qui sont concomitantes, prospectives ou rétrospectives par rapport à sa prise de décision, autrement dit sa confiance en soi ou en autrui sur la base de l'introspection des intentions personnelles et de l'évaluation des dispositions du partenaire. Ces corrélats sont proposés comme prédicteurs émotionnels ou cognitifs – selon les préférences doctrinales – de la décision prochaine du joueur dans l'incertitude sur le choix ultérieur du protagoniste. Mais on ignore toujours s'il existe un substrat cérébral des règles de proportionnalité – règles explicites comme les lois : le Code civil, et surtout règles implicites comme les régulations du marché – dans la distribution des rôles sociaux des participants à une interaction intersubjective non préméditée en un contexte de communication étendu à la société entière. Or, normalement, l'ouverture et la variabilité du contexte n'empêchent pas les personnes de prendre des décisions et d'entrer en interaction. En outre, le choix a priori des règles du jeu par l'expérimentateur au laboratoire réduit inévitablement à une certaine trivialité le caractère multidimensionnel de l'interaction intersubjective. Si le cerveau est, assurément, une direction de fondation légitime pour les interactions sociales, sauf à poser le cerveau comme sol ontologique et non seulement comme construction théorique, cette direction est non exclusive. Primordialement, l'action humaine s'appuie sur la croyance en notre pouvoir, paradoxal sans doute, d'initier des chaînes causales en contradiction avec l'ouverture infinie de la causalité physique classique comme avec la circularité de la causalité cérébrale. L'ambition

naturaliste face à cet état de choses est de remplacer le principe formel de l'autonomie rationnelle de l'agent volontaire, dont les substrats matériels restent indéterminés, par les mécanismes fonctionnels du cerveau en demandant aux neurosciences un modèle qui soit prédictif de la décision.

Mais les neurosciences sont encore loin de satisfaire à cette demande. Parce que l'équilibre entre motivation et contrôle dans le processus mental de décision et son soubassement cérébral n'a encore jamais pu être décrit en termes nomologiques, c'est-à-dire en termes de lois causales strictes, mais plutôt en termes de liaisons causales systématiquement affaiblies : corrélation, modulation, influence, sous-jacence, implication, importance, association, etc. Plutôt qu'un événement déterministe, la décision est donc ce qu'une nouvelle information pourra toujours nous contraindre à réviser, selon la formule de BAYES ; ou encore, ce dont la prévision de l'occurrence pourra éventuellement être améliorée par la connaissance des valeurs passées de séries temporelles indépendantes de l'événement en question, selon la causalité GRANGER. Une corrélation n'ayant pas l'univocité d'une vraie cause, n'exclut pas la contribution de toute autre chose que le bon sens nous recommande, à savoir les informations sur les comportements, le caractère, le milieu, l'histoire, etc. En un mot, le territoire des sciences de l'homme. Les données d'imagerie cérébrale dont on fait grand cas, ne sont ni des faits absolus comme si on regardait dans le cerveau en acte ou dans l'esprit en acte, ni nécessairement des artefacts d'ailleurs. Il n'est nullement dans mon intention de dénigrer l'imagerie cérébrale. J'entends seulement rappeler qu'elle fait apparaître certaines récurrences significatives dont il faut interroger le caractère de signifiante dans un contexte théorique plus large que le cerveau : la société, bien sûr.

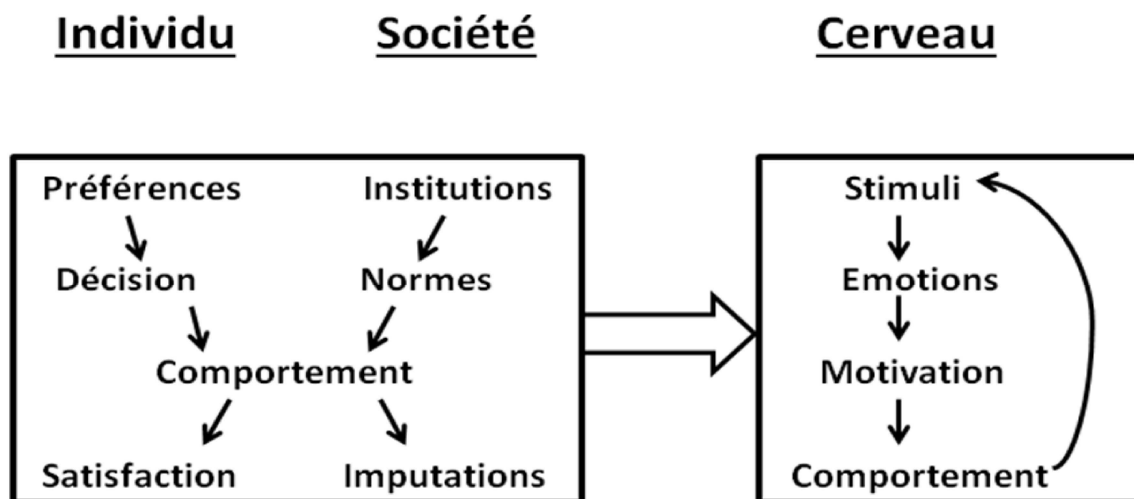
Là-dessus, permettez-moi de rappeler notre engagement de philosophes pour la rationalité de l'action. On ne peut pas abandonner complètement l'idéal de rationalité sans renoncer à comprendre l'action humaine au plan phénoménal comme au plan logique. En phénoménologie, la visée intentionnelle de l'action a un pôle objet : « la chose à faire » qui, pour être intelligible, doit être a priori structuré de façon logique, en couches de sens, autour d'un noyau d'objectivité physique support de l'intention de vérité, à la base de la valeur et de la réalisation de l'action. Je renvoie à HUSSERL : *Ideen I*¹⁸, qui est mon guide en phénoménologie. Il s'agit de fonder le sens de l'action, le sens dont on a dit ce matin qu'il disparaissait du travail en entreprise. Cela

¹⁸ Husserl E., *Idées directrices pour une phénoménologie*, trad. P. Ricœur, Gallimard, Paris, 1950 : à l'appui du caractère logique de l'action Husserl (§. 148 b) rappelle que Kant tenait la proposition « Qui veut la fin veut les moyens » pour une proposition analytique (*Fondements de la Métaphysique des mœurs* A 417).

exige une logique transcendantale à trois niveaux : ontologie formelle du jugement, axiologie de la valeur, praxéologie de l'agir. Parallèlement, en philosophie analytique, a été introduit un « principe de charité » dans l'interprétation, qui légitime la projection de la logique de l'observateur sur le comportement de l'agent observé. La même école analytique, comprenant l'insuffisance de la logique propositionnelle pour la délibération en vue de l'action, a également renoué avec la tradition du syllogisme pratique d'ARISTOTE, comme formulation d'une structure a priori encadrant notre compréhension de la décision d'autrui en tant que jugement, décision pragmatique.

Cette condition transcendantale de rationalité de l'action explique peut-être que le fonctionnement du cerveau dans la décision puisse dans la littérature neuroscientifique être couramment modélisé à l'aide des formalismes de la théorie économique la plus classique, en particulier le modèle bayésien. Non sans paradoxe : pour comprendre l'économie, nous disent les neurosciences, il faut d'abord comprendre le cerveau des décideurs économiques ; mais, pour comprendre le cerveau – système d'allocation de ressources en quantité limitée : système typiquement économique, donc – les mêmes neurosciences n'hésitent à emprunter leurs formalismes aux économistes !

Pour me résumer je poserai l'alternative : *homo economicus* ou *homonculus economicus* ?



L'individu se comprend en termes de préférences, décision, comportement, satisfaction ; la société en termes d'institutions ou normes, et à nouveau de comportements, d'imputations. Cette structure d'action individuelle, ouverte sur son contexte englobant, dont elle renoue toujours à nouveau à soi-même le réseau des relations constituantes, est-il possible de la ramener au cerveau ? Je ne dis pas mon cerveau, je dis « le » cerveau – en entendant par là

l'objet théorique des neurosciences. Le cerveau fonctionne en boucle, en circuit fermé. Les stimuli sont des modifications internes, de même les émotions, les motivations, les comportements qui changent les stimuli, et ainsi de suite.

Pour en avoir le cœur net je m'adresse aux parangons de l'individualisme économique : MORGENSTERN et VON NEUMANN. En dépit de tout leur individualisme, ils proposent des stratégies dominantes, qui sont des normes de comportements individuels, certes, mais qui garantissent la stabilité d'un *système social* dans la mesure où ces comportements règlent – ou sont censés régler en principe, sinon en fait – les problèmes de distribution de la production sociale de manière acceptable par toutes les parties.

En contraste avec cette sage reconnaissance de dette à l'égard du social de la part des économistes classiques – de qui on ne l'attendrait pas – certaines déclarations des neuroéconomistes paraîtront arrogantes, sinon impérialistes. Colin CAMERER, neuroéconomiste de la côte Ouest, estime « qu'au point de vue des neurosciences, le comportement n'est que l'un des nombreux mécanismes utilisés par le cerveau pour maintenir son homéostasie¹⁹ ». Pour un philosophe, c'est une énormité ! Quant à Paul GLIMCHER, neuroéconomiste de la côte Est, « les cartes cérébrales qui codent la valeur relative des actions possibles pour la prise de décision – dont certaines sont déjà identifiées au point de vue physiologique – sont, d'après lui, le sujet réel de la théorie économique²⁰ ».

Valeur et décision sont-elles donc biodéterminées ? L'ambition suprême de la neuroéconomie est, d'après les auteurs eux-mêmes, d'unifier les sciences de l'homme en s'appliquant à l'ensemble de nos interactions sociales dans toute leur variété. Ce que la neuroéconomie – dont, au grand jamais, je ne voudrais minimiser l'importance et les perspectives d'avenir – a montré en définitive, c'est que les bases neurales de la valeur et de la décision ne sont pas spécifiques du comportement économique, mais que l'adaptation à la vie en société complexe a imposé le remploi des mécanismes de comportements élémentaires mis en place par l'évolution pour la survie de l'organisme. La nature, disait mon maître Paul RICŒUR, nous a laissés au milieu du gué. Un fossé épistémologique reste à franchir entre les bases du métabolisme de l'organisme et une explication purement naturaliste de l'émergence

¹⁹ Camerer C., Loewenstein G., Prelec D., Neuroeconomics: how neuroscience can inform economics, *Journal of Economic Literature* XLIII, 2005, 27.

²⁰ Glimcher P., Dorris M.C., Bayer H.M., Physiological utility theory and the neuroeconomics of choice, *Games and Economic Behavior*, 52, 2005, 238.

dans le monde social de comportements individuels conformes ou non conformes aux exigences formelles normatives, qui sont celles des systèmes d'interactions humaines.

Question impertinente : est-ce qu'on complète vraiment le tableau de la décision économique en ajoutant aux conditions formelles essentielles qui font de cette décision un objet de science exacte, ses conditions mentales contingentes : confiance, regret, envie, etc., mises en avant par la neuroéconomie ? Que la réponse n'est pas nécessairement négative, je le concèderai volontiers, – à condition, toutefois, qu'on m'accorde que les formes du comportement économique ne sont pas des essences éternelles de la nature inscrites dans le cerveau humain, mais plutôt des produits de constitution intersubjective.