

ATELIER PHILOSOPHIE DE L'ACTION & NEUROSCIENCES -III

*«Modèle interne:
du bon usage d'un concept flou
en neurosciences»*

COLLEGE DE FRANCE
(11, place Marcelin Berthelot, Salle 4)

15 JUIN 2001
9h - 18h

Sous la direction de
Alain BERTHOZ

Jean-Luc PETIT
(Université Strasbourg II / LPPA)

Bernard FELTZ
(Université de Louvain-la-Neuve)

Ali BENMAKHOUF
(Université de Paris-X-Nanterre)

Jacques DROULEZ
(LPPA UMR C9950)

Joe McINTYRE
(LPPA UMR C9950)

Jean-Arcady MEYER
(LIP6)

Giuseppe LONGO
(CNRS ENS DMI)

Modèle interne : aujourd'hui, les chercheurs font volontiers appel à cette notion lorsqu'ils doivent rendre compte de certains faits que leur familiarité n'empêche pas d'être, à la réflexion, fort surprenants. Il y a, par exemple, le fait que lorsqu'on capture une balle, le cerveau n'attend pas que la balle touche la main pour produire une contraction musculaire, et que la contraction produite au moment où la balle touche la main est exactement celle qu'il faut pour contrebalancer les forces produites par la balle. La précision de cet ajustement leur a fait penser que le cerveau devait disposer d'un modèle interne des lois de la gravitation, par quoi ils entendent des réseaux neuronaux capables de *simuler* les lois physiques de l'interaction corps - environnement. Mais comme la notion de modèle implique déjà elle-même une certaine *similarité* entre les propriétés de ce modèle interne et celles du monde extérieur, on peut craindre que cette progression de concept à concept allant de «modèle interne» à «simulation», et retour, ne tombe dans un cercle vicieux.

De même, on évoquera le fait qu'on peut dessiner avec le doigt un cercle presque parfait malgré les perturbations dues à l'environnement si on le fait sur quelque chose de mou, ou d'inégalement résistant. Ce fait a donné à penser que le cerveau devait renfermer un modèle interne de la dynamique des membres, et que les propriétés physiques des os, des tendons, des muscles, et peut-être aussi les lois de la géométrie, devaient être inscrites dans ses structures nerveuses. La possession de ce modèle interne permettrait au cerveau d'*anticiper* (au sens des roboticiens) sur le mouvement, en l'esquissant de façon immanente sur ce modèle interne du bras au niveau du cervelet, avant de procéder à sa réalisation au moyen du membre effecteur lui-même dans l'espace extra-corporel. Toutefois, là encore, la notion de modèle suppléant celle de théorie dont elle garde le pouvoir *prédictif*, on peut craindre que le renvoi de «modèle interne» à «anticipation», et vice versa, ne cache un cercle vicieux.

N'imaginant cependant pas qu'une analyse logique aussi superficielle puisse suffire à disposer d'un concept en vogue chez les chercheurs, pour qui son opacité logique ne retire rien à son pouvoir de suggestion heuristique, et qui sont sensibles à des enjeux plus importants pour eux, comme la nécessité de réaffirmer la spontanéité de l'organisme, source interne de l'interprétation perceptive et de l'orientation intentionnelle des mouvements, contre le poids d'une longue tradition le réduisant tantôt à un relais sensori-moteur entre événements périphériques, tantôt à un miroir mental qui les refléterait passivement, ces motifs de perplexité n'en sont pas moins le symptôme d'une situation épistémologique qu'il devient urgent de tirer au clair. Excellente occasion de rapprocher l'investigation empirique et la réflexion philosophique sur la science dans l'examen en commun de certaines des lignes de forces qui parcourent le champ épistémologique des neurosciences actuelles. Car, le concept de modèle interne est au foyer d'un ensemble de pressions contraires, sinon contradictoires, dont on se contentera de mentionner ci-après quelques-unes des sources théoriques :

- (1) (1) (a) Une psychologie computationnelle qui présuppose un organisme cognitif aux capacités de calcul illimitées, et (b) un mouvement de cognition incarnée et située qui cherche une plus grande fidélité aux limitations effectives de l'organisme.
- (2) (2) (a) Une théorie représentationnelle de l'esprit, dispositif intériorisant des copies des objets ou matrice d'hypothèses théoriques à confirmer, et (b) une tendance à réhabiliter le rôle de l'action dans la constitution (du sens) du monde perçu plutôt que de limiter cette action au pouvoir d'y introduire des changements physiques.

(3) (3) (a) La révolution des idées (issue de la phénoménologie et de la microphysique) tendant à remplacer partout le dogme de l'objectivité absolue et de la causalité universelle par une conception de l'objectivité en constante constitution dans les interactions pratiques entre agents connaissants, et (b) une théorie causale de l'action qui retient les sciences cognitives dans l'orbite du déterminisme classique.

Eu égard à ces divers pôles d'influence théoriques, le modèle interne est, tour à tour, mais non sans tension et paradoxe :

(1) (1) mécanisme de la *neuro-computation* du cerveau-machine (de Turing) (1a), mais qui simplifie cette neuro-computation par *des processus neuro-mimétiques* localement efficaces bien que logiquement non orthodoxes (1b);
(2) (2) intériorisation des propriétés physiques des objets en *une théorie physique naïve* (2a), mais aussi *analogon concret support de manipulations internes* permettant d'anticiper sans risque les conséquences de l'action (2b);
(3) (3) mécanisme implémentant la constitution d'objets porteurs de valeurs d'utilité pratique pour l'agent (3a), mais aussi structuration des réseaux neuronaux donnée dans le cerveau, dont l'actualisation, dépourvue du caractère de temporalité constituante, n'est qu'un décours causal ordinaire : le temps de simuler (3b).

En un sens, le fait que nos anticipations dans l'action et la perception semblent se régler sur un modèle interne de la physique des choses, de la bio-mécanique du corps propre ou du répertoire des actions pratiques d'autrui semble assez convenablement pris en compte dans l'idée que l'organisme a intériorisé sous la forme de contraintes pour la microgenèse de ses mouvements et intentions un pareil schéma des propriétés des objets du monde environnant. Mais, en un autre sens, l'opération, jamais sérieusement questionnée en sciences cognitives, de l'introjection à l'organisme de ce modèle interne, et plus encore le statut proprement *métaphysique* de ce qu'on suppose comme devant être «à l'extérieur de l'organisme», «avant son intervention active», l'original de ce modèle, infiltre subrepticement dans ces sciences cognitives le dogme du monde *indépendant* des activités perceptives et pratiques de ceux pour qui il y a ce monde—et néanmoins chargé pour eux de significations.

Jacques DROULEZ

Contraintes et modèles internes comme reflets des régularités du monde.

Le flux d'informations fournies par nos sens est certes important mais notoirement insuffisant. Nous manquons toujours cruellement de données pour saisir toute la richesse des propriétés du monde et toute la complexité des interactions entre notre organisme et son environnement. La vision par exemple n'est qu'une toute petite lucarne spatio-temporelle ; elle ne peut capter qu'un nombre infime de photons, eux-mêmes les reflets très indirects et parcellaires de propriétés ou d'événements cachés qui seuls pourraient éventuellement nous intéresser.

L'objet de la psychophysique est de comprendre comment le cerveau parvient à élaborer – à reconstruire ? – une représentation cohérente de propriétés cachées mais pertinentes à partir des caractéristiques physiques observables mais « impertinentes ». Selon une conception classique de la perception, le cerveau doit donc résoudre un problème inverse : retrouver les propriétés distales supposées être les propriétés réelles du monde (au sens d'une physique sans doute naïve) qui ont donné naissance au stimulus proximal. En d'autres termes, retrouver la cause lointaine des interactions locales entre l'organisme et l'environnement, interactions locales dont seule une petite fraction est détectée par nos sens. D'une certaine façon, la génération des commandes motrices peut être conçue également comme un problème inverse. Nous souhaitons obtenir un certain résultat de notre action, comment déterminer à partir de ce but explicite la combinaison appropriée de séquences de contractions musculaires qui nous permettra de l'atteindre ?

Les problèmes inverses sont réputés difficiles à résoudre parce que mathématiquement « mal posés » : parfois il n'y a aucune bonne solution, parfois il en existe une infinité et les outils computationnels qui permettent de les obtenir sont rarement robustes. Pour pallier cette difficulté, nous pensons que le cerveau complète les données sensorielles par des contraintes ou « hypothèses internes » appliquées aux objets perçus. Bien entendu ces hypothèses internes ne sont pas conceptualisées comme telles par le sujet percevant. Elles s'expriment plutôt sous la forme « opérationnelle » de contraintes réduisant l'incertitude des données et la dimension (nombre de degrés de liberté) du problème, ou de guides simplifiant et conférant une plus grande robustesse aux processus de traitement. Ces contraintes peuvent effectivement être considérées comme des modèles internes dans le sens où elles reflètent un certain nombre de régularités du monde, parmi lesquelles on trouve évidemment certaines lois physiques générales mais aussi, plus simplement, les propriétés usuelles de notre organisme et de son environnement. Il s'agirait donc de modèles internes à valeur plutôt statistique qu'heuristique. En ce sens, la perception résulterait d'un arbitrage entre les contraintes singulières fournies par les signaux sensori-moteurs et les contraintes régulières reflétant les propriétés statistiques du monde physique observé à notre échelle : les événements perçus sont rares, les événements rares ne sont pas perçus. Je me propose d'illustrer cette idée par quelques exemples tirés de la perception visuelle du mouvement.

Joe McINTYRE

Modèles internes pour l'anticipation dans la capture d'objets en mouvement.

Comment peut-on prévoir la trajectoire d'une balle en chute libre et synchroniser l'activité musculaire avec le moment du contact ? Les études de David Lee (1981) soutiennent l'hypothèse que l'estimation du temps avant le contact se fonde sur la variable τ , ce paramètre visuel donne directement le temps avant le contact en supposant que l'objet continuera son trajet avec une vitesse constante. Cette hypothèse, issue de l'école gibsonienne : les perceptions mènent à l'action, est construite uniquement sur les signaux sensoriels que le cerveau peut mesurer directement. En revanche, Lacquaniti et al. (1989) ont montré que, lors de l'attrape d'une balle en chute libre, le système nerveux central active les muscles à un temps fixe avant le contact avec la main, indépendamment de la hauteur du lâcher. De plus, il existe une corrélation monotone entre la quantité de mouvement de la balle au moment de l'impact et l'amplitude des signaux électromyographiques anticipés. Ces deux observations suggèrent que le sujet peut prendre en compte l'accélération de la balle afin de prévoir les paramètres temporeux et énergétiques de l'impact. Comme le système nerveux central ne semble pas pouvoir mesurer l'accélération de la balle en temps réel grâce aux informations visuelles (Werkhoven et al. 1992), Lacquaniti (1993) a fait l'hypothèse que l'accélération est

prévue par une supposition *a priori* de l'action de la pesanteur terrestre sur les objets en vol. La réponse motrice (l'activité musculaire) à un stimulus sensoriel (l'image en mouvement de la balle sur la rétine) est ainsi modulée par une représentation cognitive des propriétés physiques de l'environnement. L'utilisation de cette connaissance implicite des lois physiques lors des tâches sensori-motrices est un exemple de « modèle interne ». Dans un ensemble d'expériences effectuées au sol et en apesanteur nous avons examiné les modèles internes mis en jeu par le système nerveux central lors d'une tâche de capture d'une balle en chute libre.

Giuseppe LONGO

Représentations de l'espace et du temps en mathématiques et en physique, leur rôle en cognition.

La géométrie nous propose une façon explicite de rendre intelligible, d'organiser, l'espace physique et sensible. Au cours de l'histoire, les grands tournants de cette "reconstruction" de l'espace, basée sur des invariants et les transformations qui préservent ces invariants ("l'action d'un groupe"), ont été motivés par la volonté de répondre à des questions, essentiellement les mêmes : quels sont nos observables? comment puis-je mesurer, agir sur l'espace? Or, l'analyse de nos modes d'accès au monde et de nos modes d'action sur lui est un problème typiquement cognitif. Réciproquement, les sciences cognitives portent la marque des principaux paradigmes d'investigation des fondements des mathématiques et de la physique, en dépit du fait que les courants majoritaires du domaine de la cognition continuent de se référer à une vision laplacienne de l'espace et du temps. Pour contribuer à débloquent cette situation, il faudra donc, en un double mouvement, faire en sorte qu'un dialogue se noue entre les trois grandes approches de la physique du XX^{ème} siècle, d'un côté, et les analyses contemporaines de la cognition, de l'autre. —Que peuvent apporter les différentes conceptions de l'espace (et du temps), avec leurs visions respectives de la question des fondements, à la cognition (et vice-versa)?

Jean-Luc PETIT

Modèle interne de l'action, —ou constitution kinesthésique ?

Réfléchir et discuter l'usage que diverses équipes de chercheurs en physiologie de l'action et de la perception font aujourd'hui du concept de *modèle interne* nous éclairera sur l'orientation de cette discipline particulière par rapport aux grandes tendances de la recherche contemporaine : tel est le bénéfice philosophique immédiat que nous espérons en retirer. En effet, il a été fortement suggéré que les neurosciences cognitives se trouvaient à la croisée des chemins. D'un côté, une tradition où les influences mêlées du primat du théorique par rapport au pratique, de la vision par rapport à l'action, de l'esprit par rapport au corps, du langage par rapport à l'expérience, du calcul symbolique par rapport à l'intuition géométrique, etc. se sont cristallisées sous la forme familière de la théorie de l'esprit du cognitivisme, doctrine unanimement adoptée par les psychologues et qui tend à infiltrer les disciplines voisines : physiologie et sociologie. De l'autre, l'idée d'une nouvelle physiologie de l'anticipation, forme d'intégration future d'un ensemble d'exigences récurrentes et de propositions de rechange en vue de réhabiliter, sur une base neuro-biologique, les aspects du vivant dont la phénoménologie a payé de sa mise à l'écart des courants majoritaires de la philosophie à prétention scientifique l'importance qu'elle leur a d'emblée et constamment reconnus : priorité du sens par rapport à la donnée, enracinement corporel de l'action, contribution du percevant au monde perçu, intentionnalité non linguistique de la perception, interaction

mutuelle agent - environnement, etc. Replacé dans le contexte de ces tendances contraires, le concept de *modèle interne* apparaît soumis à des tensions qui en menacent l'unité. En un sens, le fait d'intercaler un modèle interne de la biomécanique du membre effecteur, ou de la physique de l'environnement entre la composition du programme moteur et l'émission de l'ordre moteur d'une action témoigne d'une certaine reconnaissance de la spontanéité anticipatrice de l'organisme dans la conception de celle-ci (sa «simulation mentale»). Mais, en un autre sens, le concept de *modèle interne* ne diffère essentiellement ni par le sens, ni par la référence, du concept cognitiviste de représentation mentale : ce qu'on désigne *modèle interne* peut à peu près indifféremment s'appeler représentation, image, copie, carte, code, plan, schème, savoir implicite, théorie, algorithme, compétence, apprentissage, etc.; quant à *ce dont* il peut y avoir modèle interne, pratiquement tout objet possible de représentation mentale en fait partie: les objets, états de choses, événements, actions propres ou étrangères, organes du corps propre, le corps entier, l'environnement. S'il ne revient pas au philosophe de donner des normes d'usage à la recherche empirique, on ne peut lui tenir rigueur de vouloir contribuer à l'orientation générale des démarches de celle-ci par le rappel de termes de comparaison utiles dans la littérature philosophique. Dans cet esprit de modeste contribution à l'éclaircissement des idées, celle de Husserl : fonder sur l'action tout le sens du monde perçu et enraciner cette action dans les kinesthèses du sujet percevant, sera mise en regard de celle d'un modèle interne au cerveau des effecteur, instrument, objet et contexte de l'action.

Bernard FELTZ

Les paradoxes liés au concept de « modèle interne »

De nombreuses philosophies de l'esprit visent à éliminer l'idée de la conscience comme « théâtre cartésien » (D. Dennett, P. Churchland...). Un certain nombre de théories scientifiques peuvent être interprétées en un premier temps comme s'inscrivant dans cette perspective : la théorie de la sélection des groupes neuronaux de G. Edelman ou la théorie de la conscience-noyau et de la conscience étendue de A. Damasio prennent de la distance par rapport à la conception d'un centre de la conscience et y substituent l'idée d'une dynamique de réseaux. Néanmoins, avec le concept de conscience, ces théories tendent à réintroduire la logique de la représentation. Je voudrais approfondir cette contradiction apparente qui rejoint la problématique du « modèle interne », au centre de nos débats.

Jean-Arcady MEYER

Modèles internes en robotique autonome

Après avoir évoqué la controverse récente sur la nécessité de modèles internes en robotique autonome, une revue de ces modèles sera proposée et illustrée par différents exemples d'applications. On insistera, en particulier, sur les mécanismes d'acquisition de ces modèles internes - par apprentissage et/ou développement individuel, par évolution spécifique ou par une combinaison des processus adaptatifs correspondants. On évoquera également les moyens mis en œuvre pour maintenir la cohérence de ces modèles internes et pour assurer un fonctionnement correct du robot, malgré l'inévitable imprécision de ses senseurs et de ses actionneurs.

Ali BENMAKHOUF

La forme logique de la conviction

Partant de la forme logique de la conviction (Frege) nous tenterons de rapprocher celle-ci de la forme logique de la croyance (Russell) afin de préciser à l'intention des sciences cognitives ce que pourrait être un modèle interne de l'assentiment au contenu d'un jugement.