

Les Sciences Cognitives : objectivisme versus constructivisme.

1. L'objectivisme.

Le but des sciences cognitives est de constituer la cognition en tant qu'objet scientifique. Dès lors, il y a une question qui ne saurait être éludée : celle de la nature de la relation entre le sujet et l'objet de la connaissance. La position que je qualifie par le terme "objectivisme"¹ consiste à considérer que l'objet de la cognition est une réalité unique, qui existe et qui est ce qu'elle est indépendamment de toute relation (éventuelle) avec un sujet cognitif. Dans cette optique, la vérité consiste en l'adéquation entre les connaissances du sujet et cette réalité objective, indépendante et référentielle. Pour les sciences cognitives, l'objectivisme est une question épistémologique, mais en général il est aussi une question ontologique car il correspond à un engagement métaphysique concernant la nature du réel. L'objectivisme consiste à postuler que le réel est unique, qu'il préexiste d'une manière déterminée avant toute acte cognitive, et qu'il est (idéalement) connaissable en tant que tel.

Putnam (ref) a remarqué que l'objectivisme, et tout particulièrement la croyance selon laquelle la vérité consiste en l'adéquation (idéalement, un isomorphisme) entre les connaissances d'un sujet et une réalité objective indépendante, est l'une des thèses les mieux partagées dans la quasi-totalité de notre tradition philosophique occidentale depuis Platon. Encore aujourd'hui l'objectivisme est massivement majoritaire. Tout d'abord, il s'accorde avec le sens commun ; nous sommes tous, spontanément, des objectivistes dans notre vie quotidienne. Il est un fait que quand je perçois une table ou une maison (par exemple), je suis habité par la conviction que l'objet de ma perception existe et est ce qu'elle est indépendamment du fait que je porte ou non mon regard sur lui. D'autre part, cette conviction est fermement établie comme idéal régulateur régissant normativement le travail scientifique (tout au moins, comme l'a remarqué Latour (ref), pendant les périodes de stabilité des théories ; je reviendrai sur ce point). Dans notre culture, l'objectivisme prend la forme du "scientisme", et en particulier le "physicalisme", qui considère que le prototype indépassable de la connaissance scientifique est la mécanique newtonienne et le déterminisme qui y est associé. La physique possède une solide réputation, partagée non seulement par la plupart des physiciens mais, grâce à la vulgarisation scientifique, par le grand public, comme étant en droit la science de "tout ce qui est".

Historiquement, dans les sciences cognitives, le premier paradigme est celui de la théorie computationnelle de l'esprit. Cette théorie est *formaliste* : elle considère que la cognition consiste en des opérations purement syntaxiques sur des symboles formels. Puisque la connaissance scientifique est une forme de cognition, cette théorie possède comme corollaire évident la croyance selon laquelle des opérations purement syntaxiques sur des symboles formels constitue un cadre suffisant pour l'expression de toute connaissance

¹ On notera que ceux qui adoptent et/ou préconisent l'objectivisme préfèrent le terme « réalisme ». Par mon choix du terme « objectivisme » j'annonce donc d'emblée que j'adopterai une position critique.

scientifique quel qu'elle soit. Des liens profonds unissent l'objectivisme et le formalisme dans toute discipline scientifique pour des raisons épistémiques très générales. Mais ces liens sont particulièrement forts dans le cas du paradigme computationnel, en raison de la réflexivité incontournable des sciences cognitives. Ici, la cognition comme *objet* de science (et non seulement comme sujet épistémique) est identifiée avec la manipulation de représentations symboliques selon les règles d'une syntaxe formelle. Ces symboles formels n'acquièrent le statut de représentations cognitives ayant un contenu sémantique que par le biais de leurs relations de correspondance avec une réalité objective. Autrement dit, ce paradigme formaliste est nécessairement objectiviste pour des raisons liées à la cohérence de son propre objet scientifique.

Néanmoins, la situation n'est pas dépourvue d'une certaine ironie. Parmi les multiples mérites de l'objectivisme, et notamment du formalisme qui lui est associé, il y a une qui paraît essentielle à mes yeux. Dans son aboutissement au cours de la première moitié du vingtième siècle, le formalisme met irrévocablement en évidence ses propres limites *de l'intérieur de son propre cadre théorique*. Il s'agit du phénomène connu sous le vocable de la "Crise des Fondements". Au sein du formalisme en général, la crise est identifiée par les théorèmes d'incomplétude de Gödel. Le formalisme radical fut initialement inventé, par Hilbert, dans le cadre mathématique d'une interrogation sur les fondements des systèmes axiomatiques. Quand le formalisme était redéployé en dehors de ce cadre, pour constituer la théorie computationnelle de l'esprit et ainsi de fournir aux sciences cognitives leur paradigme originaire, la crise s'est considérablement aggravée car elle s'est transformée en anomalie majeure au sein d'une science empirique. En termes de l'épistémologie des sciences cognitives, cette anomalie se présente sous la forme de "l'indétermination du référent" (Putnam), de "l'opacité du référent" (Quine), ou encore du problème de "l'ancrage des symboles" (Harnad).

Je ne prétends nullement que ces difficultés sont de nature à réfuter la théorie computationnelle de l'esprit définitivement et sans appel. Néanmoins, en prenant le risque de s'incarner comme *contenu* d'une science empirique, l'objectivisme a rencontré ses limites. Au meilleur de lui-même, au moment de sa plus grande ambition, l'objectivisme s'est révélé radicalement incapable de s'auto-justifier. Autrement dit, grâce précisément à la théorie computationnelle de l'esprit, on sait dorénavant que l'on ne peut exclure la possibilité en droit d'une alternative à l'objectivisme.

Ainsi se pose, en principe tout au moins, la question d'une théorie non-objectiviste. La chose n'est pas aisée ; on peut se demander si l'objectivisme ne doit sa pérennité en grande partie au fait que la plupart des alternatives concevables pour nous sont d'une absurdité patente. Prenons par exemple le monisme idéaliste de Berkeley, que l'on peut résumer par la phrase : *esse est percipi*, "être c'est être perçu". Selon cette doctrine, une entité qui ne serait pas perçue - un arbre qui tomberait dans une forêt où il n'y a personne, ou bien (à son époque) la face cachée de la lune - n'existerait simplement pas. Cette thèse est extrêmement

provocante, et elle n'a pas manqué de provoquer. Par exemple, Samuel Johnson est réputé s'être exclamé : "Je réfute Berkeley *ainsi* ...", en donnant un coup de pied énergique dans une grosse pierre. Evidemment, loin d'une réfutation, l'anecdote constitue une magistrale illustration de la justesse de la thèse de Berkeley, car Johnson a très certainement perçue la pierre (ne serait-ce qu'en se faisant mal au pied !), donc Berkeley lui-même ne nierait nullement l'existence de la dite pierre. Plus généralement, (et plus sérieusement), la thèse de Berkeley est intéressante précisément parce qu'elle n'est pas directement réfutable. Cependant, elle est tellement contraire à l'intuition, au bon sens commun, qu'elle n'est guère acceptable comme telle. Une autre alternative concevable à laquelle on pense immédiatement est celle du relativisme ; mais celui-ci ne paraît pas seulement absurde, mais de surcroît moralement répugnant, car il semble substituer à l'idéal régulateur de la vérité la brutale loi du plus fort.

Il convient ici de dire quelques mots sur le rôle accordé à l'intuition ; en particulier, je viens d'expliquer et de justifier le refus du monisme idéaliste par le fait que cela est contraire à l'intuition. Or il est bien reconnu que l'intuition spontanée du sens commun est faillible au plus haut point. De plus, il est bien établi, depuis Bachelard au moins, que les concepts scientifiques sont acquis au prix d'un arrachement par rapport au sens commun. Un bon exemple en est fourni par la loi fondamentale de la mécanique newtonienne, $F = m.a$ (la force est proportionnelle à l'accélération). Cette loi possède comme corollaire la loi galiléenne de l'inertie du mouvement : en l'absence de toute force, un corps continuera indéfiniment son mouvement rectiligne à vitesse constante. Cette loi est totalement contraire à l'intuition ordinaire, car personne n'a jamais vu un tel mouvement. Intuitivement, on dirait plutôt que la force est proportionnelle à la vitesse, ce qui s'accorde bien mieux avec l'expérience quotidienne : pour faire avancer un véhicule, il faut pousser ! Mais si l'intuition a besoin d'être contrôlée, et à l'occasion profondément bouleversée par une "rupture épistémologique", cela ne veut pas dire qu'elle est purement et simplement bafouée. Après une rupture épistémologique - ou à l'intérieur de la science, après une "révolution" kuhnienne - l'intuition *se reconstitue*. Dans le cadre de l'exemple cité, on se dit qu'à l'intérieur d'un véhicule (par exemple, un avion ou un TGV) il est vrai que l'on ne ressent pas la vitesse ; une force ne provient, en effet, que d'une accélération ou décélération. De même, dans un mouvement circulaire, on ne ressent aucune force dans le sens de la marche, où la vitesse peut être considérable mais constante ; on ressent une force centrifuge dans une direction *perpendiculaire* au sens de la marche, où la vitesse est nulle mais où l'accélération (pour dévier d'une ligne droite) est constante. Par ailleurs, la vitesse constante d'un véhicule provient du fait que la force exercée pour le faire avancer est contrebalancée par la friction, force de résistance au mouvement. Et ainsi de suite, jusqu'au moment où l'intuition et le point de vue scientifique se rejoignent pour combler la brèche ouverte par la rupture épistémologique. Il se forme une *nouvelle* intuition dont Polanyi (1958) a remarqué, à juste titre, que les scientifiques eux-mêmes ne sauraient se passer. Le monisme éliminationniste est

inacceptable non pas parce qu'il surprend et renouvelle notre intuition, mais parce qu'il l'évacue sans rien offrir en échange.

2. Le constructivisme.

Je nommerai ici "constructivisme" le projet consistant à élaborer, dans le cadre des sciences cognitives et donc comme objet scientifique, une alternative acceptable à l'objectivisme. Les fausses solutions de l'idéalisme et du relativisme définissent au moins le cahier de charges pour une telle entreprise : le constructivisme doit rendre justice à nos intuitions concernant les contraintes d'un principe de réalité. Si l'on reprend les choses à la base, le constructivisme est bien obligé de prendre le contre-pied de l'objectivisme en niant qu'il existe une réalité objective, référentielle, unique, qui serait en droit spécifiable indépendamment de toute relation à un sujet cognitif. Autrement dit, le constructivisme se doit de considérer que le sujet et l'objet de la connaissance sont, dans un certain sens, inséparables. Mais contre l'idéalisme, le constructivisme aura à maintenir une *distinction* entre sujet et objet. En ce concerne le relativisme, le constructivisme admet qu'il peut y avoir (en droit certainement, et probablement en fait aussi) une pluralité de "réalités". Néanmoins, le constructivisme fait valoir que la condition de possibilité du venir à être de l'une ou l'autre de ces réalités est que le couple sujet-objet doit prendre une forme déterminée, et que même si le nombre de ces formes pouvant potentiellement advenir est en droit illimitée, elles ne sont pas quelconques. (Une métaphore approximative pouvait être celle des nombres premiers, qui sont tout sauf quelconques parmi les entiers, les rationnels, les réels, les nombres complexes). Par ailleurs, à *l'intérieur* de chaque forme particulière, tout n'est pas possible non plus ; on peut considérer que la "viabilité" de cette forme exige que les choses doivent se passer d'une certaine manière et pas autrement. Puisque l'on ne peut exister qu'à l'intérieur d'une forme déterminée, en fin de compte *tout se passera comme si* la cognition devait se conformer à une réalité objective extérieur (qui n'est, en fait, qu'une pure "projection" de la stabilité qui résulte de l'opération de ces doubles contraintes). Ainsi, contre le relativisme, le constructivisme se doit de fournir une explication des *apparences* objectivistes.

Il est intéressant de s'attarder quelques instants sur les relations mutuelles entre la théorie computationnelle, qui est intrinsèquement objectiviste, et le constructivisme qui se donne l'ambition d'être non-objectiviste. La tentation pour une théorie objectiviste, telle le computationnalisme, d'évincer son rival est grande. Si la recherche scientifique consiste à connaître le vrai, et si de surcroît le vrai consiste en l'adéquation entre des connaissances et une réalité objective indépendante, l'explication "sociologisante" de l'existence d'un paradigme dominant paraît superflue. Dans le cadre de l'objectivisme, le fait qu'il n'existe généralement qu'un seul paradigme à la fois s'explique par le fait que dans un domaine donné il ne peut y avoir deux théories différentes que si l'une (au moins) est fautive et donc condamnée à disparaître. De même, la stabilité temporelle des grandes théories scientifiques s'explique tout simplement par le fait qu'elles sont vraies. De ce point de vue, le discours

kuhnien sur les "révolutions scientifiques" prend l'allure d'une dramatisation excessive, d'une rhétorique abusive. Prenons par exemple les deux "révolutions" de la physique au début du vingtième siècle, celle de la relativité einsteinienne et celle de la mécanique quantique. On peut très bien faire valoir que d'un point de vue formel, la mécanique newtonienne classique est un cas particulier de la mécanique quantique (quand le constant de Planck tend vers zéro), et un cas particulier de la relativité (quand la vitesse de la lumière tend vers l'infini). Ladite "révolution" prend alors l'allure d'une généralisation qui, loin d'être incommensurable, inclut les théories précédentes comme des cas particuliers. Autrement dit, pour résumer, l'objectivisme admet difficilement un pluralisme épistémologique. Néanmoins, il est littéralement vital pour le computationnalisme de renoncer à la tentation hégémonique. J'ai déjà remarqué que l'objectivisme, et le formalisme qui lui est associé, ne peut s'auto-justifier dans l'absolu. L'affaire devient particulièrement délicate dans le cas des sciences cognitives, en raison de la réflexivité de celles-ci ; pour la théorie computationnelle de l'esprit, il y va de la validité de son propre objet scientifique. Si la théorie computationnelle essaie d'éliminer son rival constructiviste, elle n'aboutit qu'à sa propre autodestruction. L'argument est le suivant.

Selon l'objectivisme, la vérité consiste en un isomorphisme entre des connaissances scientifiques et une réalité unique. Par conséquent, s'il existe deux paradigmes qui sont réellement différents - et le computationnalisme objectiviste et le constructivisme non-objectiviste le sont assurément - l'un des deux (au moins) doit être faux ; car si les deux étaient vrais, donc isomorphes à cette même réalité unique, ils seraient nécessairement isomorphes entre eux et donc ne seraient pas vraiment différents. Ainsi, si elle veut éliminer le constructivisme, il est suffisant mais aussi nécessaire pour la théorie computationnelle de *prouver* qu'elle est vraie dans ses propres termes, à savoir qu'elle possède un isomorphisme *garanti* avec une réalité indépendante. Mais, en raison de la crise des fondements et le problème insoluble de l'indétermination du référent (ref Putnam), elle est en devoir de reconnaître que cette tâche est au-dessus de ses forces. CQFD. Il ne lui reste, pragmatiquement, que la solution de la tolérance.

Il est plaisant de constater qu'au-delà de leurs différences (qui sont fondamentales), il y a une certaine réciprocité dans cette tolérance forcée. Le constructivisme, étant par sa nature pluraliste, non seulement peut mais doit reconnaître une légitimité à son rival computationnaliste. Le constructivisme peut et doit s'efforcer de démontrer sa "vérité" dans ses propres termes, à savoir qu'il satisfait des contraintes analogues à la "viabilité" d'un organisme vivant. Mais cela ne l'autorise nullement à conclure qu'une autre forme de connaissance ne peut être "viable" aussi. Un constructivisme qui s'embarquerait imprudemment dans le projet consistant à vouloir réfuter définitivement la théorie computationnelle de l'esprit, en essayant de faire valoir que puisqu'il est lui-même vrai le computationnalisme ne peut l'être, se trahirait lui-même en réhabilitant, implicitement mais radicalement, le critère objectiviste. Autrement dit, un constructivisme cohérent ne peut que

légitimer le computationnalisme ; certes, non pas dans ses prétentions hégémoniques et absolutistes, mais dans sa version pragmatique qui est, de toute évidence, viable. Aussi paradoxale que cela puisse paraître, le constructivisme est en fin de compte mieux à même de légitimer le computationnalisme que le computationnalisme lui-même.

Cette tolérance croisée obligée, qui est propre aux sciences cognitives *quelque soit le paradigme adopté* en raison de leur projet fondamental consistant à "naturaliser l'épistémologie", ajoute une dimension supplémentaire aux considérations "sociologisantes". Cela explique peut-être, en partie tout au moins, que l'on se trouve dans cette situation assez inhabituelle où deux paradigmes, nullement complémentaires mais au contraire fortement antagonistes, peuvent néanmoins être amenés à coexister. Il n'y a aucune raison pour que cette coexistence soit particulièrement pacifique, et même beaucoup de raisons pour qu'elle ne le soit pas ; mais elle est peut-être appelée à être durable.

3. La disciplinarité.

Traditionnellement, la recherche scientifique est structurée en termes de champs dont la constitution épistémique est fournie par des *disciplines* - la physique, la chimie, la biologie, la psychologie, la linguistique, les sciences sociales etc. Chacun de ces champs est lui-même structuré par des sous-disciplines - par exemple, la zoologie, la botanique, la microbiologie, l'écologie etc. pour la biologie, ou encore la sociologie, l'ethnologie, l'économie, l'histoire etc. pour les sciences sociales. Historiquement, les "sciences cognitives" en tant que telles étaient initialement constituées par la théorie computationnelle de l'esprit qui a vu le jour dans les années 1940 dans le contexte des Conférences Macy (refs Heims, Dupuy). Mais à cette époque certains aspects de la cognition avaient déjà été étudiés par des disciplines telles la psychologie, la linguistique et les neurosciences, qui existaient au moins depuis le début du vingtième siècle. Il est donc impossible d'éluder la question du rapport entre les sciences cognitives et un certain nombre de disciplines scientifiques qui leur ont précédé et qui continuent encore aujourd'hui d'avoir une existence plus ou moins indépendante.

Que ce soit dans la guise du paradigme computationnel ou celle du paradigme constructiviste, les sciences cognitives possèdent un important point commun : elles établissent un point de vue qui est *transversal* par rapport à la multiplicité des disciplines particulières qui y contribuent. En effet, tout paradigme qui ambitionne de constituer la "cognition" comme un authentique objet scientifique se doit, sous peine d'être frappé de nullité, d'être en mesure d'articuler ces différentes disciplines entre elles. Quand on connaît les avatars de l'interdisciplinarité, si souvent préconisée mais si rarement réalisée en raison de l'énergie avec laquelle les disciplines académiques érigent des frontières et la férocité avec laquelle elles traquent les transfrontaliers, il est évident que la chose n'est pas aisée. Néanmoins, chacun à sa manière, le computationnalisme et le constructivisme réussissent tous les deux cet exploit. Comme je l'ai déjà indiqué, le premier à y parvenir historiquement était le computationnalisme ; quelque soit ses défauts (et je ne ménagerai pas mes critiques à son

égard), il faut reconnaître que sans son œuvre pionnier les sciences cognitives n'existeraient simplement pas aujourd'hui.

La question de la liste complète des disciplines qui appartiennent aux sciences cognitives est litigieuse et tout sauf claire à l'heure actuelle. En fait, aucune discipline n'appartient automatiquement aux sciences cognitives. Il est certainement possible de construire des paradigmes, dans quelque discipline que ce soit, de telle sorte que leur "pureté" disciplinaire est garantie et donc que leur appartenance aux sciences cognitives soit exclue. Par exemple, il existe des manières de faire de la neurophysiologie et de la neuroanatomie qui se limitent strictement à l'étude des molécules et des cellules individuelles, sans prendre en compte le comportement des réseaux de neurones, et l'inscription de ceux-ci dans le corps d'un animal qui lui-même agit dans le monde. Or des molécules, ou des neurones isolés (par exemple en culture *in vitro*), n'ont rien de "cognitif" en eux-mêmes ; par conséquent, ce genre de neuroscience s'exclut elle-même des sciences cognitives. Pour prendre un autre exemple, il est possible de constituer en objet scientifique nommé "langue" un système structurel qui ne renvoie qu'à lui-même ; une telle linguistique est non seulement autonome, elle est positivement autarcique d'une manière qui interdit son articulation avec la psycholinguistique et la sociolinguistique. Un tel objet scientifique, déconnecté aussi bien de son dépendance sur des processus neurophysiologiques qui le soutiennent que de ses fonctions de médiation de relations sociales et de modification de la psychologie des individus, n'a plus rien de cognitif ; par conséquent, ce genre de linguistique (parfaitement légitime par ailleurs) n'appartient pas non plus aux sciences cognitives.

En fait, la plupart des disciplines scientifiques, qu'elles appartiennent aux sciences naturelles ou aux sciences humaines et sociales, ne sont intrinsèquement ni objectivistes ni non-objectivistes. C'est bien pour cette raison que les scientifiques dans leur immense majorité peuvent considérer que l'épistémologie, et *a fortiori* l'ontologie, relèvent de considérations "métaphysiques" au sens péjoratif du terme et qu'ils peuvent très bien s'en passer. Mais dès lors qu'il s'agit d'une affiliation avec les sciences cognitives, cette "neutralité" n'est plus possible. Les sciences cognitives, elles, sont *nécessairement* soit objectivistes, soit non-objectivistes, car il y va de la constitution de leur propre objet scientifique. Comme nous l'avons vu, c'est cette particularité qui explique l'opposition *structurelle* entre les paradigmes computationnel et constructiviste. Chaque paradigme cherchera à recruter et à mobiliser le plus de disciplines possibles. Mais ce faisant, l'action conjointe de ces deux paradigmes met en lumière une *tripartition* qui passe à *l'intérieur* de toutes les disciplines scientifiques. Cela est particulièrement clair pour les disciplines qui appartiennent *a priori* aux sciences cognitives. Par exemple, il existe une linguistique autonome, qui comme nous l'avons vu n'appartient pas aux sciences cognitives et qui de ce fait peut éventuellement (mais pas nécessairement) esquiver la question de l'objectivisme ; il y a aussi une linguistique objectiviste, qui trouvera les fondements de sa sémantique dans la référence à une réalité indépendante et préexistante, et qui de ce fait est apte à être recruté par

le paradigme computationnaliste ; et il existe enfin une linguistique non-objectiviste, qui prend en charge le rôle de chaque langue dans l'avènement même d'une réalité particulière et qui, de ce fait, est apte à être mobilisée par la paradigme constructiviste. Mais cela est vrai en général, pour toutes les disciplines quelles qu'elles soient.

A cet égard, le cas de la biologie est particulièrement intéressant. Beaucoup de sous-disciplines biologiques se présentent comme n'ayant aucun parti-pris théorique particulier, mais plutôt comme une simple "description" empirique de la réalité. Malgré son apparente innocuité, cette position est en réalité profondément objectiviste ; le fait que l'objectivisme en question soit implicite plutôt qu'affiché n'y change rien. En effet, on ne peut croire qu'une description puisse être "simple" et ne requiert aucun engagement théorique de la part de l'observateur qui si on pense que la réalité est ce qu'elle est indépendamment de toute relation au sujet (collectif) de la discipline en question. En fait, étant donné la radicalité de l'opposition entre le computationnalisme objectiviste et formaliste d'une part et le constructivisme non-objectiviste de l'autre, la recombinaison des savoirs induite par l'émergence de ces deux paradigmes conduit à démasquer ce genre de fausse neutralité. Dès lors qu'une biologie "empirique" rejoint les sciences cognitives, elle est obligée d'explicitement les présupposées théoriques qui l'ont fondée épistémiquement. Si, comme la majeure partie de la biologie contemporaine, elle s'inscrit dans le cadre de la théorie "synthétique" néo-darwinienne, elle sera nécessairement formaliste (rappelons que la génétique est par excellence une discipline formelle) et objectiviste (la conceptualisation de la sélection naturelle comme mécanisme adaptatif présuppose la préexistence d'une réalité déterminée auquel il faut s'adapter); de ce fait elle formera tout naturellement une alliance avec le paradigme computationnaliste. Si, par contre, elle est mobilisée par le paradigme constructiviste, elle se trouvera dans l'obligation de se doter d'un fondement épistémique explicitement non-objectiviste.

Plus généralement, la question des disciplines qui appartiennent ou non aux sciences cognitives est aujourd'hui largement ouverte. En fait, comme on pourrait s'y attendre, la question ne se pose pas de la même manière suivant que l'on se place dans le cadre du paradigme computationnaliste ou constructiviste. La logique, notamment, occupe une place primordiale dans le computationnalisme, alors qu'elle est tout à fait subalterne pour le constructivisme. Celui-ci, par contre, sollicite fortement les disciplines ayant pour objet la genèse des organismes vivants - la paléontologie, l'étude comparative des espèces, la préhistoire et l'histoire ; et à une autre échelle de temps, l'embryologie. En psychologie, le constructivisme est proche de l'Ecole du Gestalt et la psychologie développementale, alors que le computationnalisme est à l'origine même du courant dit "psychologie cognitive". Par ailleurs, les deux paradigmes mobilisent très différemment les sciences sociales. Le computationnalisme privilégie les disciplines formalisées - l'économie, la théorie des choix rationnels, et dans la sociologie l'individualisme méthodologique. Le constructivisme mobilise plutôt l'anthropologie, l'ethnologie, l'ethnométhodologie, et en sociologie la tradition

herméneutique et la sociologie de la connaissance. Enfin, les deux paradigmes entretiennent des relations très différentes avec la philosophie. Le computationnalisme est même issu de la tradition anglo-saxonne de la philosophie analytique, où l'empirisme logique joue un rôle déterminant. Le constructivisme, par contre, est en affinité avec la tradition sceptique (ref von Glasersfeld) et, plus récemment, la phénoménologie.

4. La transduction sujet-objet.

4.1. Introduction.

Reprenons donc la question de l'objectivisme. Comme je l'ai déjà indiqué, il paraît à première vue complètement insensé d'imaginer que l'existence d'un objet (de la cognition) puisse ne pas être indépendante de sa relation avec un sujet cognitif. Cela semble nous amener tout droit à la provocation absurde de Berkeley : "ce qui n'est pas perçu n'existe pas". Imaginons par exemple le continent américain. En oubliant ou tenant pour quantité négligeable les amérindiens autochtones (comme le veut notre "civilisation" européenne), avant l'arrivée de Christophe Colomb l'Amérique n'était pas perçu. Est-ce que cela veut dire qu'elle n'existait pas, mais qu'au moment précis où Colomb pointe à l'horizon elle surgit soudain de nulle part ? Non, bien sûr. Mais j'ai dit que la thèse de Berkeley n'est pas réfutable, et il sera intéressant à présent de voir pourquoi. En effet, afin de réfuter Berkeley, il faudrait mettre en évidence de manière concluante l'existence d'un objet non-perçu. Mais alors, le piège se referme ; car une telle mise en évidence, si elle est pleinement convaincante, constitue en elle-même une forme de perception. La faille dans l'exemple de Colomb et l'Amérique provient de ce que nous avons oublié notre propre existence en tant qu'observateurs discrets mais omniscients. En imaginant la scène, *nous* avons perçu l'Amérique (ne serait-ce que par reconstruction rationnelle) avant que Colomb ne franchisse l'horizon.

Retenons donc notre agacement, et prenons le problème par l'autre bout. Y a-t-il certains objets, ne serait-ce que quelques uns à titre illustratif, dont il n'est *pas* absurde de dire qu'ils n'existent pas en l'absence d'un sujet cognitif ? Prenons par exemple les odeurs. En l'absence d'un sujet doté de capacités olfactives, il y a peut-être des molécules chimiques (je reviendrai sur ce point), mais il n'y a pas *d'odeurs* en tant que telles. L'exemple est d'autant plus intéressant qu'avec les avancées récentes dans nos connaissances concernant le dispositif sensoriel de l'olfaction - les molécules réceptrices mais aussi les réseaux neurophysiologiques qui sont activés lors de l'olfaction - il devient de plus en plus clair que ce qui constitue une odeur saillante dépend fortement des caractéristiques du sujet de la perception (ref Freeman). Un autre exemple, tout à fait analogue, est celui des couleurs. En l'absence d'un sujet cognitif, il y a peut-être des ondes électromagnétiques dans une certaine gamme de longueur d'onde, mais il n'y a pas de *couleurs* proprement dites. Le physicalisme réductionniste le dit bien, depuis longtemps, en faisant une distinction entre des qualités "primaires" (les masses, charges électriques, positions et vitesses des particules élémentaires) qui seules sont

pleinement "réelles", et les qualités "secondaires" (couleurs et formes visuelles, sons auditifs, toucher, odeurs et goûts - il est à remarquer que celles-ci correspondent très précisément au "cinq sens") qui sont en quelque sorte des "citoyens de deuxième classe". Je soulève ici, par anticipation, une question que j'aborderai de manière plus conséquente par la suite. La primauté accordée par les physiciens aux "qualités primaires" ne provient-il pas de leur propre particularité *en tant qu'observateurs* ? Eux ne perçoivent que les qualités primaires ; mais qu'est-ce qui pourrait justifier que l'on accorde un privilège *ontologique* à leur point de vue ? Mais revenons pour l'instant aux questions soulevées par des objets, tels les odeurs et les couleurs, dont il n'est manifestement pas absurde de dire qu'ils n'existent pas en l'absence d'un sujet cognitif.

Je propose de nommer "transduction" ce type de relation entre deux entités où chacun des pôles est *constituée* par sa relation à l'autre (ref Simondon). Si l'un des pôles vient à manquer, la relation disparaît et, avec elle, l'autre pôle disparaît aussi. Mais en même temps, les deux pôles sont distincts, sans quoi il n'y aurait pas de *relation*. En philosophie, l'étude de ce type de relation entre sujet et objet a été menée par la phénoménologie, dont le grand initiateur était Husserl et l'auteur le plus important pour nous est sans doute Merleau-Ponty. Dans les termes techniques de la phénoménologie, on parle de la *visée intentionnelle*. L'objet d'une visée intentionnelle ne doit en aucun cas être confondu avec un (éventuel) "objet réel là-bas dans le monde". La phénoménologie évite soigneusement de se prononcer sur l'existence ou la non-existence de cette "réalité objective", qui est "mise entre parenthèse" par une opération techniquement nommée *l'épochè*. Il est à souligner que cette conception n'est nullement idéaliste ; la "visée intentionnelle" peut être plus ou moins parfaitement (voire pas du tout) "remplie". Par rapport à une odeur, par exemple, on peut dire "ça me rappelle un peu de l'herbe coupé, mais ce n'est pas tout à fait ça", ou au contraire "c'est *exactement* ça". Mais en même temps, cette conception n'est pas objectiviste : la constitution d'une visée intentionnelle dépend manifestement du sujet. Par exemple, les oenologues, après un long et patient travail aboutissant à la construction d'un répertoire cognitif d'une grande richesse, sont capables d'identifier immédiatement des odeurs très précises là où, pour la plupart d'entre nous, il n'y a qu'un vague arôme. Pourtant, nous sommes tous en présence des mêmes molécules.

4.2. La couleur.

Il sera utile d'examiner dans un peu plus de détail l'exemple d'un "objet transductif" qu'est la couleur. D'une part, des avancées récentes dans la neurophysiologie de la vision rendent beaucoup plus concrète l'idée selon laquelle le sujet participe au venir à être des objets de sa perception. D'autre part, comme nous allons le voir par la suite, cet exemple de la couleur a joué un rôle déterminant dans l'élaboration de la pensée de Maturana et Varela qui sera d'une importance capitale pour mon propos.

Les cellules rétiniennes qui sont nécessaires pour la vision des couleurs sont maintenant bien identifiées. Chez les êtres humains, il existe trois sortes de cellules "coniques", ayant chacun un pigment photosensible à une longueur d'onde relativement spécifique : on parle des récepteurs d'ondes longues (L), d'ondes moyennes (M) et d'ondes courtes (C). En effet, les courbes d'absorption spectrale se chevauchent, mais atteignent respectivement leurs sommets à 560, 530 et 440 nanomètres. Mais les couleurs perçues ne sont pas un reflet simple de l'activation de ces récepteurs. Les neurones dans la partie du cerveau concernée interagissent fortement entre eux par des processus tantôt excitants, tantôt inhibiteurs, de tel sorte que l'ensemble forme un "réseau" fortement organisé et ayant des propriétés émergentes particulières. Plus précisément, cette organisation neurophysiologique donne lieu à trois "canaux". Le premier canal, appelé "rouge-vert", est constitué par la *différence* entre les signaux provenant des récepteurs L et M. Le deuxième canal, "bleu-jaune", est constitué par la différence entre la somme des signaux provenant des récepteurs L et M et les signaux des récepteurs C ($L + M - C$). Chacun de ces canaux "chromatiques" est "opposant", c'est-à-dire qu'une augmentation du rouge est toujours acquise au détriment du vert, et vice versa ; et une augmentation du bleu est toujours acquise au détriment du jaune, et inversement. Ainsi s'explique le fait, dont tous les êtres humains normaux (à l'exception, par exemple, des daltoniens chez qui le canal rouge-vert est déficient) peuvent témoigner, qu'il n'est pas possible de percevoir *à la fois* du rouge et du vert, ni du bleu *et* du jaune. Par contre, les deux canaux chromatiques peuvent se combiner, dans des proportions variables. Ainsi s'explique le fait que, pour nous, il est possible de passer progressivement du rouge (canal 1) au bleu (canal 2), en passant par des violets plus ou moins rougeâtres ou bleutés, sans solution de continuité ; du rouge au jaune en passant par des oranges ; du vert au bleu en passant par des turquoises ; et du vert au jaune en passant par des verts jaunâtres et des jaunes verdâtres. Le troisième canal, celui de la luminosité, est constitué par la somme des signaux provenant des trois récepteurs : selon cette dimension, les couleurs varient de l'aveuglant au pâle. Autrement dit, l'ensemble de ce que sont les couleurs pour nous en tant que "visées intentionnelles" est caractérisé par une structure phénoménologique qui est à la fois précise et, à bien réfléchir, très particulière. Il y a lieu de penser que pour d'autres animaux, ayant d'autres photo-pigments et d'autres organisations neurophysiologiques, "les couleurs" seraient toute autre chose que pour nous êtres humains. Varela (ref), à qui j'ai emprunté l'essentiel de cette explication concernant la perception des couleurs, emploie le terme "énaction" pour désigner la relation transductive entre objet et sujet de la perception, et cite fort à propos un passage de Merleau-Ponty :

"Ainsi la forme de l'excitant est *créée par* l'organisme lui-même, par sa manière propre de s'offrir aux actions du dehors. Sans doute, pour pouvoir subsister, il doit rencontrer autour de lui un certain nombre d'agents physiques et chimiques. Mais c'est lui, selon la nature propre de ses récepteurs, selon les seuils de ses centres nerveux, selon les mouvements des organes, *qui choisit dans le monde physique les stimuli auxquels il*

sera sensible Ce serait un clavier qui se meut lui-même, de manière à offrir... telles ou telles de ses notes à l'action en elle-même monotone d'un marteau extérieur."

4.3. Perception et action.

Il semble bien, donc, que certains objets comme les couleurs sont transductifs car inséparables de la structuration du sujet. Mais on peut à la fois consolider, étendre et radicaliser cette conclusion en faisant valoir que sans une *action* de la part du sujet, la perception en tant que telle *n'existe simplement pas*. Autrement dit, la simple activation d'un organe sensorielle ne suffit pas pour donner lieu à une perception. Je propose d'illustrer ce point en me référant principalement à une remarquable série d'expériences menées par Paul Bach y Rita (refs). Ces expériences datent initialement des années 1970 ; largement négligées pendant 20 ans, elles connaissent actuellement un important regain d'intérêt. Le dispositif expérimental, appelé de façon quelque peu trompeuse "substitution sensorielle", est le suivant. La sortie d'une petite vidéo-caméra est transformée électroniquement pour activer une plaque ayant une matrice de 20 x 20 points de stimulation tactile d'intensité variable. Cette plaque est fixée sur la poitrine du sujet (aveugle, ou éventuellement voyant mais ayant les yeux bandés), de sorte que l'image optique de la caméra est "projetée" pour produire une sensation tactile correspondante. Mais les sujets ne "voient" pas immédiatement pour autant ; comme on vient de le dire, il faut "apprendre" à voir. Dans le premier protocole d'apprentissage, qui servira en fait de "contrôle", l'expérimentateur informe le sujet oralement sur l'état de l'image : "maintenant, le champ est uniformément éclairé" (le sujet ressent un picotement sur l'ensemble de la surface de sa peau stimulé par la plaque) ; "maintenant, c'est la moitié gauche du champ qui est éclairé", "maintenant, la moitié droite", "maintenant, il y a un triangle lumineux", etc. Les résultats de ce protocole sont franchement décevants : au bout de quelques semaines d'apprentissage quotidien, le sujet parvient péniblement à reconnaître quelques formes grossières. Corrélativement, l'expérience phénoménologique reste celle d'un picotement tactile au niveau de l'interface sensorielle, identifiée comme la surface de la peau. Tout change avec le deuxième protocole d'apprentissage, où le sujet prend lui-même la vidéo-caméra dans ses mains. Il n'y a plus besoin de l'expérimentateur pour informer le sujet, car celui-ci sait lui-même quels sont les actions qu'il a effectué pour modifier la stimulation sensorielle : s'il tourne la caméra vers la gauche, "l'image" (indifféremment "optique" ou "tactile") va se déplacer vers la droite ; s'il tourne la caméra vers le haut, l'image se déplace de façon correspondante, etc. L'effet est encore amplifié si le sujet dispose d'autres actions ayant des effets en retour sur ses sensations, par exemple en activant un bouton de "zoom". Avec ce deuxième protocole, les résultats sont spectaculaires ; au bout de quelques heures seulement, le sujet "perçoit" clairement des formes, et au bout de quelques jours parvient à des performances de reconnaissance remarquables. Par exemple, entendant un léger bruit, il tourne la caméra vers la porte où quelqu'un vient d'entrer, et dit : "Marie, pourquoi ne t'es-tu pas coiffé ce matin ?" Autre anecdote révélateur : si l'expérimentateur, à l'insu du sujet,

augmente brusquement le "zoom", le sujet produit un mouvement reflexe de recul en se protégeant la figure avec son bras : en effet, "normalement" cette modification de l'image serait produit par un grand objet se rapprochant brusquement (et dangereusement) du sujet.

L'interprétation de cette expérimentation est claire : à la différence de la simple sensation (que l'on peut définir comme la stimulation passive d'un organe sensoriel, comme dans le premier protocole), la *perception* proprement dite requiert la participation *active* du sujet. Autrement dit, la perception est fondamentalement une perception des conséquences des actions pour les sensations. En effet, c'est sur ces bases que le sujet peut former une perception d'un monde extérieur stable, qui correspond précisément à ce qui reste invariant dans le couplage sensori-moteur. C'est sur ce "fond" que le sujet peut faire des distinctions "figure-fond", et distinguer les modifications de l'état de ses organes sensoriels qui proviennent de ses propres actions (le fond) et celles qui n'en proviennent pas et qui seront donc identifiés comme des objets *extérieurs* qui bougent. On voit bien à quel point les objets de la perception sont des objets *transductifs*, inséparables de l'activité du sujet lui-même. Dans les expériences de Bach y Rita, cette interprétation est consolidée et enrichie par les corrélats phénoménologiques. En effet, j'ai déjà indiqué que dans les conditions du premier protocole, l'expérience phénoménologique reste celle d'un picotement tactile à surface de la peau. Dans les conditions du deuxième protocole, il y a un changement radical : le sujet n'identifie plus le "lieu" de la perception comme étant la surface de sa peau : les objets perçus sont *projetés* "là-bas dans le monde". Du coup, sa conscience du picotement à la surface de sa peau s'efface - de même que, lors de la vision normale, nous n'avons aucune conscience d'une sensation au niveau de nos yeux ; nous percevons les objets "directement" comme existant "là-bas". Bien sûr, il est possible pour le sujet de refocaliser sa conscience du picotement cutané (de même qu'une intensité douloureuse, quand nous regardons le soleil directement par exemple, peut raviver une conscience de la participation de nos yeux) mais alors, c'est notre conscience des objets de la perception qui s'efface à son tour. Ce phénomène de "projection" des objets perçus n'est d'ailleurs nullement spécifique à la vision, et peut facilement être expérimenté par tout un chacun. Si je ferme les yeux, et je "tâtonne" avec une baguette ou même un stylo tenu à la main (ce que je fais à l'instant, en écrivant ces lignes), je perçois les objets sur le bureau (l'ordinateur, le clavier, le carnet, le téléphone....) non pas au bout de mes doigts (là où il y a les terminaisons nerveuses les plus proches des objets), mais *au bout du stylo*. De même, quand on conduit une voiture, les objets que l'on perçoit ne se situent pas au niveau des mains (sur le volant), des pieds (sur les pédales du frein et de l'embrayage), et encore moins celui des cuisses et des fesses ; ils sont projetés au point de contact entre la route et les pneus, qui risquent de glisser ou de heurter une bosse.

L'identification d'un "monde extérieur stable" avec un invariant sensori-moteur éclaire aussi un fait important que nous ne remarquons généralement pas car nous le prenons (à tort) comme allant de soi. Quand nous tournons notre tête, ce qui provoque un déplacement de l'image rétinienne, nous n'avons pas l'impression que c'est "le monde" qui tourne. La non-

trivialité de ce phénomène est mise en évidence par des expériences (très dangereuses, et malheureusement exploitables et exploitées à des fins de torture) de privation sensorielle. En effet, après une dizaine de jours de privation sensorielle totale (ce qui abolit les retours des conséquences des actions pour les sensations), les repères que sont ces invariants sensori-moteurs se perdent, avec la conséquence que quand le sujet tourne sa tête il a effectivement l'impression que c'est "le monde qui tourne", ce qui provoque vertiges et désorientation. Le fait que la perception corresponde à l'identification d'invariants sensori-moteurs explique aussi pourquoi, bien que l'image rétinienne soit "à l'envers", nous voyons le monde "à l'endroit" ; et explique aussi pourquoi, quelques jours après avoir endossé des lunettes qui font voir le monde "à l'envers" (en mettant l'image rétinienne "à l'endroit" !), les sujets se remettent à voir le monde "à l'endroit". Finalement, dans le même ordre d'idées, il s'avère que des modifications de l'activation des cellules rétiniennes à la suite des actions du sujet sont une condition nécessaire à la "sensation" elle-même. En effet, si les muscles qui produisent les mouvements saccadés des yeux sont paralysés, ou si (plus subtilement encore) un dispositif optique est installé qui compensent les effets des mouvements saccadés de sorte que l'image rétinienne n'est pas altérée, *on ne voit plus* un point lumineux qui continue pourtant de stimuler les cellules rétiniennes.

4.4. Les mondes animaux.

Il semble donc légitime de conclure que les objets de la perception sont profondément transductifs, c'est-à-dire n'existent que dans leur relation à un sujet. Je propose de laisser ouverte, pour l'instant, la question à savoir si *tous* les objets cognitifs, quels qu'ils soient, sont transductifs. J'ai déjà évoqué la question soulevé par le physicalisme réductionniste, qui accorde une primauté ontologique aux "qualités primaires" aux dépens de "qualités secondaires". Nous aurons, le moment venu, à soumettre la science elle-même à un regard critique, et de voir ce que deviennent les objets scientifiques dans une optique non-objectiviste. Mais ne brûlons pas les étapes ; pour l'instant, j'espère seulement avoir fait valoir qu'il existe au moins *certaines* objets qui sont transductifs ; et donc qu'une position non-objectiviste est beaucoup plus plausible qu'il ne paraît à première vue. Je souhaite à présent poursuivre dans la voie d'une naturalisation constructiviste de l'esprit en examinant s'il est possible de maintenir cette conception d'objets cognitifs transductifs en laissant de côté, provisoirement, l'accès à ces phénomènes par la voie privilégiée d'une perspective à la première personne. L'enjeu est important, car la perspective à la première personne ne nous est accessible (et encore...!) pour chacun d'entre nous individuellement et, peut-être, (par la communication linguistique) pour une communauté d'êtres humains.

Il existe une transition naturelle du monde humain vers celui des animaux, car les expériences de Bach y Rita possèdent un analogue assez précis : il s'agit des expériences désormais classiques de Held et Hein (ref). On sait que si des chatons sont élevés dans le noir, ils restent aveugles et que s'ils sont brusquement introduit à la lumière du jour ils ne peuvent

voir. De même que pour des êtres humains aveugles de naissance à qui on enlève une cataracte à l'âge adulte, il faut *apprendre* à voir. Mais Held et Hein ont montré qu'une simple exposition à la lumière ne suffit pas. Le dispositif expérimental est le suivant. Deux chatons sont mis chacun dans un harnais, les deux harnais étant reliés par une barre tournante. L'environnement est symétrique, de sorte que la stimulation visuelle reçue par chaque chaton est pratiquement identique. Il y a cependant une différence entre les situations des deux chatons. Le harnais de l'un des chatons est relevé de quelques centimètres, de sorte que ses pattes ne touchent pas le sol ; par conséquent, les mouvements de son corps ne produisent pas de modifications correspondantes dans ses sensations visuelles. Le harnais de l'autre chaton, par contre, est placé de telle sorte qu'il est en contact efficace avec le sol, et chacun de ses mouvements produit effectivement une modification en retour de ses sensations visuelles. Les deux chatons sont élevés tous les deux dans le noir, sauf pendant la période de quelques heures par jour où ils sont placés dans le dispositif. Le résultat de cette expérience est saisissant : le chaton "actif" voit normalement, alors que le chaton qui ne subit que passivement les *mêmes* stimulations optiques est presque aussi aveugle que s'il avait été élevé entièrement dans le noir. Il semble légitime de conclure que tout autant pour les animaux que pour les êtres humains, la perception proprement dite ne se construit pas sans des retours de l'action sur la sensation.

Mais on peut pousser encore plus loin, en essayant d'imaginer ce que peut être le monde *pour* des animaux non-humains. Il faut souligner qu'il ne s'agit pas d'un exercice purement spéculatif et incontrôlable : de même que pour la phénoménologie humaine, les "variations eidétiques" de Husserl sont tout sauf le fruit d'une introspection spontanée et gratuite, la construction d'une représentation des "mondes animaux" est un exercice systématique et discipliné qui s'appuie en l'occurrence sur de véritables expérimentations concernant les capacités perceptives des sujets en question. Je noterai en passant, sans entrer dans les détails, qu'un exercice de ce type est possible concernant les nourrissons qui, bien que des êtres humains en puissance, ne peuvent témoigner linguistiquement de la qualité de leur expérience vécue. Il faut lire, à ce titre, la remarquable reconstitution du "Journal d'un Bébé" (de la naissance à l'âge de quatre ans, où le petit des hommes rejoint notre monde adulte) réalisée par le psychologue Daniel Stern (ref). Mais venons-en au grand oeuvre classique qui a frayé la voie d'une exploration des "mondes animaux". L'auteur en question est l'éthologiste allemand Jacob von Uexküll, dont la lecture a beaucoup inspiré Merleau-Ponty.

Dans son livre intitulé "Mondes animaux et monde humain", von Uexküll (ref) montre qu'il est possible de caractériser ce qu'est le monde *pour* un animal non-humain, même si nous ne pouvons évidemment pas connaître ce monde par une expérience vécue directe. Son exemple le plus connu est celui du "monde de la tique". Je ne peux résister à reprendre la présentation qu'en fait von Uexküll, témoin d'une autre époque où les biologistes savaient encore écrire :

"L'habitant de la campagne qui parcourt souvent bois et buissons avec son chien, n'a pas manqué de faire connaissance avec une bête minuscule, qui, suspendue aux tiges des buissons, guette sa proie, homme ou bête, pour se précipiter sur sa victime et se gorger de son sang. La bestiole, qui n'a qu'un ou deux millimètres, se gonfle alors jusqu'à prendre la dimension d'un petit pois..... Sa vie a été si bien étudiée que nous pouvons en tracer une image sans grande lacune..... Lorsque la femelle a été fécondée, elle grimpe à l'aide de ses huit pattes jusqu'à la pointe d'une branche d'un buisson quelconque pour pouvoir, d'une hauteur suffisante, se laisser tomber sur les petits mammifères qui passent ou se faire accrocher par les animaux les plus grands."

De notre point de vue d'observateurs humains, il y a cependant un problème. Il est évident que le hasard qui fait passer un mammifère sous la branche où se trouve la tique est extraordinairement rare. Dans le monde qui entoure la tique, il y a beaucoup d'évènements - le jour et la nuit, les vents, la pluie, parfois des bruits, etc. Comment la tique peut-elle "guetter" en permanence, et se laisser tomber au moment précis du passage d'un mammifère ? La résolution de cet énigme est due à une remarquable complémentarité entre des observations de terrain et des expériences de laboratoire. D'une part, la tique possède dans une proportion inhabituelle la faculté de vivre longtemps sans nourriture ; en laboratoire, on a maintenu en vie des tiques qui avaient jeûné pendant dix-huit ans. Pendant toute cette période, la tique reste dans un état d'animation suspendue qui, tout comme le sommeil ou le coma chez nous, interrompt le passage du temps. On notera, en particulier, que les cellules séminales restent abritées chez la femelle dans des capsules. Mais le problème n'est qu'à moitié résolu ; car comment la tique peut-elle sortir de cet état d'animation suspendue au moment où passe un mammifère ? La réponse provient d'autres études de laboratoire : la tique possède un sens d'odorat très particulier qui est activé spécifiquement par l'acide butyrique. Cette observation prend tout son sens quand on sait que dans la niche écologique naturelle de la tique, une odeur d'acide butyrique est dégagée par tous les mammifères (en provenance de leurs follicules sébacés), mais uniquement par des mammifères. Nous, observateurs humains, sommes tentés d'admirer la pertinence fantastique qui permet à la tique de "faire le tri" parmi tous les stimuli potentiels provenant de son milieu et d'identifier précisément ceux qui indiquent la présence d'un mammifère. Mais quand on connaît le mécanisme, on voit que notre jugement de "pertinence" est en fait un pur artefact de *notre* connaissance de la situation. La tique, elle, n'a pas de tri à faire : elle ne perçoit *que* l'acide butyrique.

Mais le processus ne s'arrête pas là. L'activation de l'organe sensorielle déclenche une *action*, à savoir le relâchement des pattes, avec comme conséquence la chute de la tique. Si elle tombe effectivement sur un mammifère, son sens tactile est activé et déclenche un mouvement d'exploration jusqu'à ce que la tique parvienne à un endroit dépourvu de poils qu'elle commence à perforer. Des essais de laboratoire montrent que la tique n'a pas le sens du goût ; en effet, après perforation de la membrane, elle absorbe tout liquide qui a la bonne

température. Si la tique, stimulée par l'acide butyrique, tombe sur un corps froid, elle a manqué sa proie et doit regrimper à son poste d'observation.

Ce qu'il convient de noter, c'est que ces différentes capacités de sensation et d'action sont *organisées* : chaque sensation déclenche une action qui conduit à l'extinction de la sensation précédente et à l'activation de la sensation suivante.... qui déclenche une autre action, et ainsi de suite. Il y a donc une *boucle* entre perception et action que von Uexküll nomme "cercle fonctionnel". Il est à souligner que les objets du milieu participent à la constitution du cercle, mais d'une manière particulière : "l'objet ne participe à l'action qu'en tant qu'il doit posséder les caractères nécessaires qui peuvent servir d'une part comme porteurs de caractères actifs, d'autre part comme porteurs de caractères perceptifs, lesquels doivent être un connexion structurale les uns avec les autres. Les rapports de sujet à objet ressortent très clairement du schéma du cercle fonctionnel. Il montre comment le sujet et l'objet sont ajustés l'un à l'autre et forment un ensemble ordonné." On ne saurait espérer une meilleure formulation de ce que j'ai appelé les "objets transductifs".

Reprenons la citation de von Uexküll :

"La tique reste suspendue sans mouvement à une pointe de branche dans une clairière. Sa position lui offre la possibilité de tomber sur un mammifère qui viendrait à passer. De tout l'entourage aucune excitation ne lui parvient. Mais voilà que s'approche un mammifère dont le sang est indispensable à la procréation de ses descendants.

C'est à ce moment que se produit quelque chose d'étonnant.... Comment ce brigand de grand chemin, aveugle et sourd..... peut-il percevoir l'approche de sa proie ?.... De tous les effets dégagés par le corps du mammifère, il n'y en a que trois, et dans un certain ordre, qui deviennent des excitations. Dans le monde gigantesque qui entoure la tique, trois stimulants brillent comme des signaux lumineux dans les ténèbres et lui servent de poteaux indicateurs qui la conduiront au but sans défaillance. Pour cela, la tique est pourvue, en dehors de son corps avec ses récepteurs et ses effecteurs, de trois signaux perceptifs qu'elle peut transformer en caractères perceptifs. Et le déroulement des actes de la tique est si fortement prescrit par ces caractères perceptifs qu'elle ne peut produire que des caractères actifs bien déterminés.

La richesse du monde qui entoure la tique disparaît et se réduit ainsi à une forme pauvre qui consiste pour l'essentiel en trois caractères perceptifs et trois caractères actifs - son *Umwelt*. Mais la pauvreté de l'*Umwelt* conditionne la sûreté de l'action, et la sûreté est plus importante que la richesse."

C'est ici que von Uexküll introduit sa grande notion théorique, pour laquelle il utilise le néologisme allemand *Umwelt* et que l'on peut traduire en français par *monde propre*. Le "monde propre" de l'animal se définit par contraste avec le "monde en général" tel qu'il peut être perçu par nous en tant qu'observateurs humains. (Je reviendrai plus tard à la question de savoir si le "monde scientifique" n'est pas aussi un "monde propre", particulier aux êtres humains qui s'engagent dans une activité sociale spécifique). Le "monde propre" correspond,

comme je l'ai déjà indiqué, à ce qu'est "le monde" *pour* l'animal. La nature transductive du "monde propre" est, j'espère, manifeste : il est différent pour chaque espèce d'animal, et inséparable des spécificités de celle-ci.

Je terminerai cette évocation du monde propre de la tique en racontant la fin de l'histoire. L'enchaînement des trois cercles fonctionnels conduit la tique à "s'enfoncer jusqu'à la tête dans le tissu cutané (du mammifère). Elle aspire alors lentement à elle un flux de sang chaud.... Quand le sang ... parvient dans l'estomac de la tique, les cellules séminales abritées dans les capsules se libèrent et vont féconder les œufs qui seront chargés de réserves. Mais le copieux repas de sang de la tique est aussi son festin de mort, car il ne lui reste alors plus rien à faire qu'à se laisser tomber sur le sol, y déposer ses œufs et mourir".

Von Uexküll souligne, de manière exemplaire, la circularité de la relation entre perception et action. On notera, toutefois, que selon von Uexküll le "retour" de l'action sur la perception est *négatif* : "le caractère actif *éteint* le caractère perceptif" (c'est moi qui souligne). Son point de vue est donc complémentaire (mais non contradictoire) à celui qui se dégage des travaux de Bach y Rita et de Held et Hein, où nous avons vu que dans un registre temporel de la synchronicité immédiate (de l'ordre d'une fraction de seconde), les effets "en retour" de l'action sont positivement nécessaires pour que l'activation d'un organe sensoriel devienne une véritable perception. Le point de vue de von Uexküll élargit cette première perspective en montrant que dans un registre temporel plus long (de l'ordre de minutes, voire (pour nous!!) des heures ou même des années) c'est *l'enchaînement* des différents "cercles fonctionnels" qui *ensemble* constituent le "monde propre" de l'animal.

Pour compléter ce tableau des "mondes animaux", je souhaite mentionner l'apport d'un auteur qui, autant que von Uexküll, a fait une contribution majeure à l'élaboration d'une conception de la cognition des animaux selon laquelle les objets de leur cognition sont inséparables de leur propre activité. Il s'agit du psychologue américain J.J. Gibson et ce qu'il appelle une approche "écologique" de la perception. Comme von Uexküll et Bach y Rita, Gibson souligne que les animaux sont "cognitifs" non seulement parce qu'ils utilisent leurs sensations pour guider leurs actions, mais aussi et surtout parce que ce sont leurs propres actions qui *constituent* les objets possibles de leur cognition. Commençons par un exemple à la fois simple et fondamental. Quand un animal ayant des yeux se meut, il génère *du fait même de son mouvement* ce que Gibson appelle un "flux optique". Le point droit devant est la source immobile de ce flux, qui se déploie sur la surface de la "sphère" dont il est le centre pour atteindre sa valeur maximal au niveau du grand cercle orthogonal à la direction du mouvement avant de se rétrécir et finalement disparaître englouti dans le point immobile directement derrière l'animal. Ce qui est perçu par l'animal ne peut être décrite par l'optique de la physique (isotrope, rectiligne...); il ne peut être décrit que dans les termes de ce flux optique "écologique" qui est produit par l'animal lui-même. A ce propos, il est intéressant de noter qu'une mouche, par exemple (qui ne peut produire des mouvements saccadés de ses yeux composés) est littéralement aveugle à tout ce qui ne bouge pas. En effet, elle ne perçoit

que ce qui bouge, soit du fait de son propre mouvement (dans quel cas c'est bien le "flux optique" gibsonien qu'elle perçoit), soit parce que c'est un objet extérieur bouge (une main qui cherche à l'écraser, par exemple...). Qui plus est, la mouche semble bien faire la distinction entre les deux cas. Ces phénomènes ont été bien étudiés, tant sur le plan du comportement que celui des mécanismes neurophysiologiques sous-jacents (refs Campan, Franceschini).

Gibson parle donc d'une perception "directe", qu'il convient d'interpréter comme signifiant que la perception n'est pas une "représentation" d'une réalité externe (le référent objectiviste), mais le résultat direct du couplage sensori-moteur de l'animal avec son environnement qui n'a aucun besoin d'être "re-présenté" une deuxième fois. On peut faire le rapprochement ici avec le point de vue du roboticien Rodney Brooks, qui considère (non pas abstraitement, mais concrètement en fabriquant des robots) que si on parvient à mettre en place des mécanismes sensori-moteurs appropriés, on n'a plus besoin de "représentations". Brooks exprime cela en disant que "le monde est sa propre 'représentation' ". L'expression "perception directe" est quelque peu malheureuse, car (comme l'a remarqué Varela - ref) elle se prête à une interprétation diamétralement opposée, à savoir un objectivisme particulièrement naïf. Il me semble, cependant, que l'interprétation objectiviste des concepts gibsoniens n'est pas tenable ; c'est ce que nous allons voir maintenant en examinant le deuxième concept majeur de Gibson qu'il désigne par le terme "affordance".

Le mot "affordance" est un néologisme en anglais que l'on peut traduire par la périphrase "occasion offerte à l'action". Le mieux est sans doute de prendre quelques exemples. Ainsi, un arbre "offre la possibilité" de s'y percher et d'y construire un nid (si on est un oiseau), de grimper dessus (si on est un écureuil), de se mettre à l'ombre (si on est une vache par une grande chaleur d'été), de s'y fourrer pour s'abriter et se nourrir (si on est un certain type d'insecte), et ainsi de suite. Gibson déploie des effets de langue efficaces pour évoquer les *qualias* des "affordances" : "la fleur dit 'sente-moi', la pomme dit 'croque-moi', la femme dit 'aime moi' (!)". Mais je souhaite donner deux exemples supplémentaires qui illustrent bien la nature nébuleuse et insaisissable de la "réalité en soi" si l'on cherche à l'identifier de manière objectiviste sans référence à un sujet particulier. Mon premier exemple est l'air. Qu'est-ce que l'air ? Il est, littéralement, insaisissable, et plutôt incolore et invisible. Bien sûr, on peut l'identifier dans les termes de la physique (comme gaz) et de la chimie (par sa composition moléculaire). Mais ne pressent-on pas (pour anticiper sur la question des objets scientifiques) qu'il s'agit là de caractérisations assez particulières, qui en outre n'épuisent nullement ce que l'air peut "être" dès que l'on introduit des sujets particuliers ? Jugez-en vous-mêmes : si on est un oiseau, l'air *devient* le support pour une activité de vol. Et pour être encore plus précis, le *type* de support dépend du *type* d'oiseau : ce n'est pas la même chose pour un grand oiseau planeur (tels les aigles ou les buses) que pour des oiseaux comme les hirondelles qui ont un vol dynamiquement instable de façon à pouvoir changer brusquement de direction (ce que ne peuvent faire les planeurs) afin d'happer des insectes au vol. Et ce qu' "est" l'air est loin de s'arrêter là. Pour les animaux terrestres, il est le support

pour une activité de respiration ; pour les poissons, il est plutôt irrespirable ; et pour les bactéries et algues anaérobiques, il s'agit carrément d'un poison qui est le résultat de leurs propres activités de photosynthèse devenu "pollution" (pour eux) quand, au bout de 3000 millions d'années, la teneur d'oxygène est montée brusquement de quelques pourcent à sa valeur actuelle de 20% environ il y a 1000 millions d'années maintenant. Pour les plantes, en revanche, l'air est une source de nourriture sous forme de CO₂ dont elles ont besoin pour constituer matériellement leurs tissus. Et on peut sans difficulté continuer : pour des animaux ayant des capacités auditifs, l'air est le support de la transmission d'ondes sonores ; pour des animaux ayant une vision en couleurs, c'est l'air encore qui fait que le ciel est bleu (et, si on est astronaute, que la Terre est "la planète bleu"). Bref, j'espère avoir montré qu'en fin de compte, "l'air" est pratiquement autant de choses différentes qu'il existe des espèces ; et même cette liste, qui s'énumère par milliards, n'épuise pas en droit ce qu'est "l'air" *en soi*, car elle ne tient pas compte des nouvelles espèces (ou des nouveaux mondes que les êtres humains peuvent amener à exister par leurs réalisations technologiques) qui *pourraient* exister. Ce que l'air "en soi" *peut* être est littéralement illimité et inépuisable.

L'exemple de l'air n'est en rien un cas unique ; on pourrait répéter l'exercice en parlant de l'eau. On remarquera (pour contrer le dérive objectiviste de certains gibsoniens) que plus on s'attache à l'explicitation concrète de ce qu'est une "affordance" précise, plus il devient clair qu'elle est très radicalement inséparable de la stratégie de vie déployée par une "forme de vie" *particulière* (en gros, une espèce biologique, ou dans le cas des êtres humains une ethnie - ref Leroi-Gourhan). Les "mondes animaux" sont bels et bien des objets transductifs.

4.5. Cognition = Vie = Autopoïèse.

Les travaux de Varela, Bach y Rita, von Uexküll et Gibson rendent scientifiquement crédible l'idée selon laquelle non seulement des êtres humains, mais aussi des animaux même très simples, peuvent être les sujets cognitifs d'objets transductifs. Peut-on aller encore plus loin dans la généralisation, et considérer que la cognition est une faculté propre à tous les organismes vivants, même les organismes unicellulaires comme les bactéries qui sont les organismes les plus simples que nous connaissons ? Ce n'est pas évident ; von Uexküll et Gibson, notamment, considèrent qu'à la différence des animaux, les plantes n'ont pas de "monde propre". Cependant, c'est cette voie radicale que je propose d'explorer, en me référant dans un premier temps au biologiste chilien Humberto Maturana. Pour entrer sur cette voie, nous aurons besoin de poser ouvertement la question : qu'est que la vie ?

Depuis son plus jeune âge, le biologiste chilien s'est posé inlassablement la question : quelle est la caractéristique *essentielle* des organismes vivants ? Que doit être un organisme vivant pour qu'il puisse mourir ? L'approche habituelle à ce genre de question consiste à partir d'une définition de sens commun - de considérer qu'au fond on sait déjà suffisamment ce qu'est un organisme vivant, assez en tout cas pour dire sans hésitation qu'un chien est vivant alors qu'une pierre ne l'est pas - et d'examiner "empiriquement" les propriétés communes à

tous les entités ainsi catégorisées comme "vivantes". On peut avoir quelques doutes *a priori* sur le bien-fondé d'une telle démarche, si l'on se rappelle avec Bachelard que la connaissance scientifique se constitue en *rupture épistémologique* d'avec le sens commun, et avec Kant (et Kuhn) que les objets scientifiques sont donnés *en théorie* et non pas par induction à partir de "simples" observations empiriques. A tout le moins, on peut remarquer que la position (assez fréquent chez des scientifiques, et tout particulièrement chez des biologistes du vingtième siècle) selon laquelle des considérations "théoriques" seraient superflues car la science consiste à "décrire la réalité telle qu'elle est" est implicitement mais d'autant plus profondément objectiviste. En effet, on ne peut considérer que le choix d'une théorie (ou d'un paradigme) n'influera pas sur les connaissances produites que si on considère qu'il existe effectivement une réalité de référence, pré-existante et unique, et que les connaissances vraies entretiennent une relation de correspondance isomorphe avec cette réalité - ce qui est la définition même de ce que j'appelle "objectivisme". L'empirisme qui dédaigne la dimension théorique peut donc être compatible avec la paradigme computationnaliste, mais en aucun cas avec ce qui nous intéresse ici, à savoir une alternative non-objectiviste au computationnalisme.

Par ailleurs, indépendamment de ces considérations épistémologiques, on peut examiner plus pragmatiquement les résultats d'une approche empiriste à la définition du vivant. Maturana (ref) raconte comment, pendant de longues années au début de sa quête (et notamment en essayant de répondre aux questions de ses étudiants), il était obligé d'accepter que l'on pouvait reconnaître des systèmes vivants quand on les rencontrait, mais que l'on ne pouvait pas encore dire ce qu'ils étaient. Je le cite :

"Je pouvais énumérer des traits des systèmes vivants, tels que la reproduction, l'hérédité, l'irritabilité, la croissance, et ainsi de suite ; mais jusqu'où devait-on continuer la liste ? Comment pourrait-on savoir si elle était complète ? Afin de savoir quand la liste serait complète il me fallait savoir ce qu'était un organisme vivant, ce qui me ramenait à la même question qui avait motivé l'élaboration de la liste. Je pourrais parler de l'adaptation et de l'évolution, du développement et de la différenciation, et montrer comment tous ces phénomènes étaient interconnectés par le phénomène de la sélection naturelle ; mais la question "Quel est le trait invariant des systèmes vivants autour duquel la sélection naturelle opère" restait sans réponse. Chacune de ces tentatives me ramenait au point de départ".

On dirait Alice au Pays des Merveilles, cherchant à atteindre le sommet de la colline mais se retrouvant à chaque fois en train de rentrer dans la maison ! On peut rajouter que si la voie de "la liste" ne peut aboutir, l'approfondissement d'un seul élément sur la liste rencontre le même problème. Prenons, par exemple, le premier trait sur la liste de Maturana (et pas seulement de lui) : la reproduction. Première objection : les mules (par exemple) ne se reproduisent pas, mais est-ce que cela veut dire qu'ils ne sont pas vivants ? Mais cette objection n'est pas très grave ; après tout, "c'est l'exception qui confirme la règle" et les

organismes vivants qui ne se reproduisent pas sont bien les exceptions. Bien plus profonde est la *même* objection que celle qui invalide l'approche par une liste : si on ne sait pas *déjà* ce qu'est un organisme vivant, le fait que c'est une entité qui "se reproduit" ne nous renseigne pas davantage. Par exemple, dans certaines conditions des cristaux - et aujourd'hui, en pleine épidémie de la maladie des vaches folles, les prions - se "reproduisent" ; s'ensuit-il vraiment qu'ils sont "vivants" ?

A la suite de ces longues interrogations, Maturana s'est rendu compte qu'il fallait changer radicalement d'approche. Cependant, à la différence d'Alice, la réponse n'était pas immédiate. C'est petit à petit qu'il était amené à penser qu'il fallait caractériser des systèmes vivants non pas par référence à leur environnement ou leur contexte, mais par rapport à eux-mêmes en tant qu'entités *autonomes*. En 1969 il parlait pour la première fois de systèmes vivants en disant qu'ils étaient constitués en tant qu'entités par la *circularité* des processus de production de leurs propres composants. En effet, une fois qu'on y pense, cette idée possède une évidence intuitive. Si on demande : qu'est-ce qui *fabrique* un organisme vivant, il est clair que c'est l'organisme lui-même. Que ce soit un animal, une plante ou un micro-organisme, les tissus et les organes sont le résultat d'un processus dynamique permanent de *production* ; les molécules qui les composent se renouvellent continuellement. Et cela n'est vrai que pour des êtres vivants ; toute machine fabriquée par des êtres humains, par exemple - que ce soit une usine entière, une machine-outil ou autre - produit *autre chose* qu'elle-même, et en même temps elle est elle-même produite *par* autre chose qu'elle-même. Cette circularité "auto-référentielle" semble donc bien être une caractéristique essentielle des organismes vivants.

Par ailleurs, Maturana travaillait sur la neurophysiologie de la vision. Dans ses premiers travaux, en collaboration avec Lettvin, McCulloch et Pitts, il prenait comme allant de soi que la situation était clairement définie : qu'il existait une réalité objective, indépendant et externe à l'animal ; et que l'animal pouvait percevoir cette réalité et utiliser l'information acquise afin de calculer un comportement approprié. Autrement dit, sa position spontanée était ce que j'ai appelé "objectiviste". Mais à partir de 1964, quand il commençait avec Frenk et Uribe à étudier la vision des couleurs, il est rapidement devenu clair que cette approche n'allait pas pouvoir aboutir. Nous avons déjà vu que les couleurs sont des objets "transductifs", inséparables du sujet qui les perçoit. Mais il convient à présent de pousser encore plus loin cette question, en examinant de plus près la difficulté de faire correspondre une couleur *telle qu'elle est perçu* avec un "réfèrent" externe. Maturana, Frank et Uribe ont montré très clairement qu'il y a une multitude de configurations externes, ayant des compositions spectrales et des formes géométriques extrêmement variées, qui toutes donnaient lieu à une même perception. Ce point est tellement important qu'il mérite que l'on en donne un exemple.

Soit un écran, illuminé par deux projecteurs, l'un qui donne une lumière blanche, l'autre ayant un filtre qui produit une lumière rouge (voir Figure....). La couleur perçu sur

l'écran est une rouge claire ; pas de surprise pour le moment. Maintenant, on met un objet (par exemple, la main) dans le faisceau provenant du projecteur "blanc", ce qui produit une ombre sur l'écran. La couleur perçue de cette ombre est rouge foncé. Toujours pas de surprise. Mais maintenant, on met l'objet dans le faisceau "rouge" ; il y aura, évidemment, une ombre sur l'écran. Mais de quelle couleur ? Il faut le faire pour y croire ; mais les observateurs humains (non-daltoniens!) sont unanimes sur la réponse, l'ombre est *verte* ! Ce n'est pas un vert vif, couleur herbe fraîche ou d'une pomme "granny", c'est un vert grisâtre et terne, mais l'ombre est indéniablement verte. Ce résultat est assez radicalement inexplicable si on essaie de rattacher la couleur perçue à des données externes, car le fait qu'il y ait ou non un éclairage rouge supplémentaire autour ne change rigoureusement rien à la lumière reflétée de la partie de l'écran à l'abri ombragée de cette lumière ; cette lumière possède la composition spectrale de la lumière blanche, et tout instrument de mesure physique ne pouvait que confirmer que rien n'a changé. Cependant, quand il y a seulement la lumière du projecteur "blanc", le sujet perçoit du blanc ; quand il s'agit d'une ombre sur fond rouge clair, la *même* lumière est perçue comme verte.

Une variante de cette expérience confirme encore plus dramatiquement la conclusion selon laquelle les couleurs perçues ne sont pas référables à une "réalité externe". Si on place dans chacun des projecteurs des copies identiques d'un damier de gris, blancs et noirs, si les deux damiers sont exactement alignés, ce que l'on voit sur l'écran est un tableau de roses de différents degrés de saturation. Si on tourne l'un des damiers de 90 degrés, le résultat est surprenant à l'extrême : on perçoit une image multicolore, contenant de petits carrés jaunes, bleus et verts aussi bien que rouges et roses (Maturana, Uribe et Frenk 1968).

Par ailleurs, au niveau neurophysiologique, Maturana et ses collaborateurs ont démontré que l'activité du système nerveux ne pouvait pas non plus être mise en corrélation avec des "couleurs externes" définies par leurs propriétés physiques. En effet, afin de spécifier des distinctions de *couleur*, le cerveau employait des neurones dont on savait qu'ils étaient sensibles à la *géométrie* des objets visuels. Résumons les relations entre trois entités ; (i) la couleur perçue, (ii) l'activité du système nerveux, et (iii) l'état de la "réalité externe". Il n'existe pas de relation identifiable entre (i) et (iii), ni entre (ii) et (iii). Comment alors identifier la signification de (i) et (ii), si ni l'un ni l'autre est référible à la réalité externe ? La réponse, après-coup, semble si évidente que l'on s'étonne de ne pas y avoir pensé avant : la couleur perçue, (i), est corrélée avec (ii), l'activité du système nerveux.

Cette découverte, très simple en elle-même, constitue pourtant un renversement radical par rapport à nos idées habituelles. Une fois envisagée, elle s'impose à notre entendement ; en effet, comment pourrait-il en être autrement ?! Mais alors, si l'activité du système nerveux (et par conséquent la couleur perçue) résulte de la dynamique *autonome* du système nerveux lui-même, les conséquences en sont profondes. D'abord, cela signifie qu'une perception ne correspond pas à la *détection* de traits (et encore moins d'objets) externes ; elle correspond plutôt à la *spécification* par le sujet lui-même de ce qu'il y a *pour lui* dans le

monde. Ensuite, cela signifie aussi qu'il n'y a rien, et qu'il ne peut rien y avoir, dans la qualité immédiate, instantanée, d'une perception qui permet de différencier une "hallucination", une "illusion" et une "perception véridique". Il faut préciser immédiatement, pour prévenir un malentendu possible, que cette position n'est ni idéaliste, ni solipsiste, ni relativiste ; il est clair qu'en tant qu'êtres humains nous pouvons faire, et en fait nous faisons couramment, des distinctions entre des "hallucinations" et des "perceptions". Nous aurons à revenir longuement sur cette question, notamment concernant la relation entre théories scientifiques "vraies" et "fausses" ; retenons pour l'instant seulement que ce genre de distinction est moins simple qu'il n'y paraît à première vue.

Au fondement de cette nouvelle manière de voir, il y a donc l'autonomie du système nerveux. Autrement dit, le système nerveux est un système dynamique dont l'évolution temporelle est déterminée par la structure et l'organisation du système lui-même. Il interagit avec son environnement, évidemment, par les organes sensorielles et motrices ; mais ces interactions ne dictent pas l'état du système nerveux, elles sont plutôt des *perturbations* dont les conséquences dépendent du système lui-même. La question habituelle : "comment l'organisme obtient-il de l'information concernant son environnement" doit être rejetée, et remplacée par une autre : "comment se trouve-t-il que l'organisme possède une structure qui lui permet de fonctionner de manière adéquate dans le milieu où il existe ?" Si l'on veut rendre intelligible les phénomènes effectifs ayant lieu dans la perception, il faut se baser sur le fait que le système nerveux est un réseau *clos*, caractérisé par sa *circularité* auto-référentielle.

Résumons donc les conclusions essentielles de notre exploration de la possibilité de généraliser une "naturalisation de l'intentionnalité" à l'ensemble des organismes vivants. Nous avons vu que Maturana est arrivé à la conclusion que la caractéristique essentielle des organismes vivants, *qua* vivant, réside dans la circularité de leur organisation. Nous avons vu que dans un domaine apparemment différent, Maturana est arrivé à la conclusion que la perception cognitive est également caractérisée par la circularité de son organisation. D'où l'idée - à la fois simple et profonde, de l'étoffe dont sont faits les paradigmes - que "la vie" et "la cognition" sont au fond *le même phénomène*. En collaboration avec Francisco Varela, cherchant une formulation plus adéquate de la notion de "organisation circulaire", Maturana a inventé le mot "autopoïèse", du grec *autos* (soi) et *poiein* (produire). La définition canonique est la suivante (Varela 1989) :

"Un système autopoïétique est organisé comme un réseau de processus de production de composants qui (a) régénèrent continuellement par leurs transformations et leurs interactions le réseau qui les a produits, et qui (b) constituent le système en tant qu'unité concrète dans l'espace où il existe, en spécifiant le domaine topologique où il se réalise comme réseau.

Il s'ensuit qu'une machine autopoïétique engendre et spécifie continuellement sa propre organisation. Elle accomplit ce processus incessant de remplacement de ses composants, parce qu'elle est continuellement soumise à des perturbations externes, et constamment

forcée de compenser ces perturbations. Ainsi, une machine autopoïétique est un système ... à relations stables dont l'invariant fondamental est sa propre organisation (le réseau de relations qui la définit)."

La quête pour une théorie non-formaliste et non-objectiviste de la cognition semble donc pouvoir trouver un aboutissement dans une nouvelle biologie qui se donne comme objet central "la vie" en tant que telle.