

Revue européenne
des sciences sociales

European Journal of Social Sciences

Revue européenne des sciences sociales

European Journal of Social Sciences

XXXIX-121 | 2001

L'acteur. Un concept sur la scène des sciences
sociales

Une fourmi sur la plage. Remarques sur les concepts d'agent et de comportement

Henri Volken



Édition électronique

URL : <http://journals.openedition.org/ress/650>

DOI : 10.4000/ress.650

ISSN : 1663-4446

Éditeur

Librairie Droz

Édition imprimée

Date de publication : 1 novembre 2001

Pagination : 95-110

ISBN : 2-600-00663-X

ISSN : 0048-8046

Référence électronique

Henri Volken, « Une fourmi sur la plage. Remarques sur les concepts d'agent et de comportement », *Revue européenne des sciences sociales* [En ligne], XXXIX-121 | 2001, mis en ligne le 11 décembre 2009, consulté le 19 avril 2019. URL : <http://journals.openedition.org/ress/650> ; DOI : 10.4000/ress.650

Henri VOLKEN

UNE FOURMI SUR LA PLAGE REMARQUES SUR LES CONCEPTS D'AGENT ET DE COMPORTEMENT

An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself.

H. A. Simon [10]

(Une fourmi considérée comme un système comportemental, est un être relativement simple. L'apparente complexité des évolutions de son comportement est pour une grande part le reflet de la complexité de l'environnement dans lequel elle se trouve.)

A man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself.

H. A. Simon [10]

(Un homme, considéré comme un système comportemental, est un être relativement simple. L'apparente complexité des évolutions de son comportement est pour une grande part le reflet de la complexité de l'environnement dans lequel il se trouve.)

INTRODUCTION

Dans son étude sur les principes de base du comportement¹, Herbert A. Simon introduit une métaphore animale pour illustrer le point de vue qu'il va défendre. Ce point de vue peut être décrit, de manière un peu provocatrice, comme la justification du passage de la première à la deuxième phrase en exergue du présent article. Simon répond à la question : D'où vient la complexité apparente d'un comportement ? Avant de situer le problème dans un contexte plus précis, il suggère dans son livre de commencer par imaginer une fourmi se déplaçant sur une plage. Il poursuit alors sa métaphore :

Observons cette fourmi poursuivant laborieusement son chemin à travers une grève modelée par le vent et les vagues. Elle avance tout droit, tourne à droite pour escalader plus facilement une petite dune abrupte, contourne un caillou, s'arrête un instant pour échanger quelques informations avec un compatriote. Elle fraye ainsi, d'un pas hésitant, son chemin jusque chez elle. Pour éviter tout anthropomorphisme prématuré quant à ses intentions, je dessine son itinéraire sur une feuille de papier : une séquence de segments angulaires qui ne sont pourtant pas disposés au hasard ; on reconnaît un sens sous-jacent de direction, de volonté vers un but.

¹ Il s'agit du chapitre sur la psychologie de la pensée dans [10] (cf. bibliographie).

Ce dessin, l'enregistrement de la trajectoire de la fourmi, est un reflet du *comportement* de la fourmi. Mais comment ce reflet va-t-il être perçu? Comment va-t-il être interprété? Peut-il être expliqué? Est-il prévisible? Soumis tel quel et sans commentaire à différentes personnes, ce dessin provoquera des réactions diverses et probablement des réponses variées à ces questions. Certains pourront voir dans ce dessin la trace d'un grimpeur habile dans une paroi difficile et accidentée, ou celle d'un bouchon flottant sur un torrent sauvage, d'autres y verront la folle course d'un joueur de rugby portant le ballon vers le but adverse ou encore l'enregistrement du cours d'une action soumise aux aléas de la bourse.

Mais quelle que soit l'interprétation donnée à ce dessin, les deux constatations élémentaires suivantes seront toujours présentes :

- la trajectoire suggère une ligne plus ou moins droite reliant deux points
- le comportement sous-jacent, évoqué par le dessin, semble être d'une grande complexité et imprédictible, certainement hors de portée d'une explication précise

En réalité, la fourmi ne suit probablement que quelques règles très simples², lui permettant d'éviter, de surmonter ou de contourner les obstacles sans perdre de vue la direction générale suivie. En imaginant la fourmi dix fois plus grande, mais identique pour le reste et dans le même contexte, la trajectoire serait, avec certitude, beaucoup plus lisse, et une fourmi cent fois plus grande se déplacerait ici en une ligne droite presque parfaite ! La raison de cette différence d'appréciation du comportement est simple : l'environnement, c'est-à-dire la configuration de la plage, n'est pas le même pour des fourmis de tailles différentes. La complexité du comportement semble donc clairement être une fonction de l'échelle de l'acteur par rapport à la granularité de l'environnement. Une fonction, non seulement de l'échelle, mais aussi de la spécificité et des particularités accidentelles, et également des propriétés morphologiques de la fourmi et donc, plus globalement, de la confrontation sujet/environnement. En toute généralité se pose alors les questions : à qui ou à quoi faut-il attribuer la spécificité d'un comportement? Comment une explication peut-elle rendre compte d'un comportement? Et surtout, peut-on le prévoir?

La complexité du comportement, qu'on aurait naturellement tendance à expliquer par la complexité du sujet – de l'*agent*, pour anticiper notre terminologie à venir – exige ainsi un élargissement de ce point de vue comme le laisse entrevoir le simple exemple de la fourmi. La thèse de Simon, implicitement rencontrée en exergue de ce texte, est que la complexité du comportement reflète – aussi – la complexité de l'environnement. On peut aller un peu plus loin et affirmer que *le comportement n'est rien d'autre qu'une émergence de l'interaction sujet/environnement*. Ce qui signifie – nous allons discuter ce point un peu plus loin – que la complexité du comportement n'est *lisible* ni dans la fourmi ni dans l'environnement séparément. Un sujet, même de nature très simple³, peut donc se compor-

² Rien n'est dit ici sur la réalisation biologique de ces règles.

³ Certains organismes très primitifs en sont des exemples.

ter, selon l'environnement, de manière complexe et imprévisible ou au contraire de façon sage et prédictible. Par sujet – ou acteur ou encore agent – nous entendons par la suite, selon les cas, un être biologique, un homme aussi bien qu'un animal, ou une construction artificielle comme un robot ou un programme informatique. Mais ce concept polysémique d'agent ou de sujet, exige qu'on s'y attarde un peu. Nous allons tenter de préciser en particulier deux aspects importants de sa définition : un sujet est un *individu*, et cet individu possède une certaine *autonomie*.

OBJETS

L'exemple de la fourmi nous rappelle ce principe élémentaire : nous ne pouvons tenter d'appréhender le monde qu'à travers une représentation que nous nous faisons de lui. Et cette représentation elle-même ne peut se faire qu'en analysant la réalité en entités distinctes, fourmi, galets, grains de sable, végétaux etc. C'est ce découpage de la réalité en unités discrètes qui permet la perception du monde extérieur et le traitement cognitif des informations ainsi obtenues. Ces unités singulières, auxquelles nous attribuons une forme de cohérence, d'individualité, de globalité et de persistance dans le temps, nous les appellerons tout simplement *individus* ou mieux, *objets*. Pour les isoler, nous leur conférons des caractéristiques propres, mais aussi un nom qui les singularise. En résumé : un objet possède des propriétés descriptibles en termes généraux, mais aussi une identité propre qui permet de les désigner de manière univoque. Si nous lui attribuons une identité, cela signifie dans le même temps qu'il possède également une certaine pérennité.

La définition d'un objet en termes généraux suggère l'existence d'une catégorie dont celui-ci serait un spécimen. Un spécimen peut donc différer d'un autre de la même catégorie par la modalité de ses propriétés. Ainsi une fourmi peut être grande ou petite, rouge ou noire, ailée ou aptère. On peut plus généralement considérer que chaque objet est situé dans l'espace abstrait de ses caractéristiques possibles, c'est-à-dire qu'à chaque objet on peut associer son *espace des états*. Cet espace est en quelque sorte l'espace logique, l'espace de tous les objets possibles de la même catégorie.

Le discours scientifique porte la plupart du temps sur un *système*, considéré comme objet individuel d'une investigation. Ce système peut être lui-même découpé en objets plus petits, considérés individuellement à leur tour comme objets, et ainsi de suite. L'identité des objets dépend du regard que l'observateur porte sur une portion de la réalité. Par exemple on peut regarder le fonctionnement global d'un parlement et y voir deux entités enrôlées dans un processus de prise de décision : la chambre du peuple et le Sénat. Mais à l'intérieur de chaque chambre on peut aussi s'intéresser à des coalitions ou groupes de pressions, qui deviennent à leur tour des « individus ». Et, à un niveau inférieur, on peut s'intéresser dans une coalition aux membres de celle-ci, aux personnes des députés, qui accèdent par cette opération eux aussi au statut d'objets. Cette fonction de « zoom mental » correspond – mais dans le sens inverse – à l'activité du mathématicien lorsqu'il passe d'un *ensemble* à l'*ensemble de ses parties*.

Cela signifie que la notion d'environnement est elle-même fonction du point de vue de l'observateur. A travers un certain objectif, l'environnement peut très bien contenir de manière diffuse des fragments de réalité qui, à travers un autre

objectif, seront eux-mêmes distingués comme objets. Ainsi nous verrons plus tard, qu'un ensemble d'agents peut constituer un environnement pour un agent isolé.

Un objet est en général *situé* dans un système englobant et se trouve en relation avec d'autres individus. Situé signifie qu'il en est une partie intégrante et interagit avec lui à travers les liens qui le relie aux autres parties du système. Mais ces liens restent cependant suffisamment faibles pour que le caractère monadique⁴ de l'objet subsiste.

Cette attitude qui consiste à découper le monde en entités individuelles est commune à la vaste majorité des philosophes et des scientifiques. Il existe quelques rares positions philosophiques divergentes qui proposent de ne voir dans le monde qui nous entoure qu'un seul individu. C'est la position par exemple de Parménide pour qui tout ce qui est forme une seule entité, l'être. Car, selon lui, pour distinguer des parties, il faudrait, que celles-ci soient séparées par quelque chose de nature différente, donc par quelque chose qui n'est pas, ce qui est exclu. Bien entendu cette position holistique reste marginale.

Une entité de type objet peut d'ailleurs jouer le double rôle d'individu et de catégorie. C'est le cas de l'espèce en biologie. La confusion de ces deux rôles est à l'origine de nombreux dialogues de sourds dans les innombrables disputes autour de la théorie darwinienne de l'évolution.

Comment s'opère en pratique la discrimination du monde en objets? Une partie arbitraire du monde ne se présente pas forcément comme objet. Pour avoir ce statut, il doit correspondre à un concept. Bien sûr, ce concept et la catégorie qui lui est associée peuvent être plus ou moins *naturels* en ce sens qu'il reflètent une division spatiale, temporelle, géométrique, biologique ou physique du monde qui soit immédiatement perceptible. D'autres notions d'objets peuvent être plus arbitraires ce qui rend alors les concepts correspondants purement *nominaux*.

Quoi qu'il en soit, le système des objets choisis correspond à notre découpage conceptuel arbitraire et exprime notre vision du monde. Or ces entités individuelles, les objets, ne sont pas (toujours) indépendantes dans le système qui les englobe. Il sont pris la plupart du temps dans un tissu de relations causales qui les relie, et plus ce tissu est dense, plus les objets paraissent *naturels*. Sans liens entre eux, ils sont *arbitraires*. Si le concept sous-jacent est abstrait, l'objet est *idéal*. C'est le cas des objets mathématiques. Pour terminer ce petit exercice de définition, nous allons introduire le terme suivant qui est important pour la suite: un objet est dit *situé*, s'il est une partie interactive du système.

AGENTS

Si l'objet est un fragment individuel du monde avec les propriétés que nous venons de voir, *l'agent* est un objet particulier qui se caractérise essentiellement par son *autonomie* et son *adaptativité*. Par souci de simplicité nous allons com-

⁴ Voir à ce propos la théorie des monades de Leibniz.

mencer par définir l'agent tel qu'il apparaît dans les sciences cognitives et en intelligence artificielle.

D'après tout ce qui précède, il est clair que c'est le regard de l'observateur qui crée l'objet, donc aussi l'agent. Ou du moins qui le découpe sur l'arrière fond de la réalité brute. Mais qu'est-ce exactement qu'un agent ?

La particularité essentielle qui distingue un agent d'un objet individuel quelconque, nous l'avons dit plus haut, est son pouvoir de s'adapter à son environnement et une forme d'indépendance qui lui permet de prendre des initiatives. Il est doué d'autonomie. Ce dernier terme suggère qu'il crée, ou suit, sa propre loi. Qu'est-ce que cela veut dire ? D'une part qu'il est capable d'agir de manière individuelle, sans contrôle extérieur, d'autre part qu'il peut réagir, à sa manière, face à un environnement et face à d'autres agents : il est *actif*, *réactif* et *social*. Nous allons également le supposer *situé* dans le sens défini plus haut. Il possède en conséquence une faculté d'initiative et d'imprévisibilité qui le détache du réseau causal strict de son environnement. Pour un observateur extérieur, il est donc à l'origine d'un comportement qu'on ne peut pas prévoir entièrement à partir du système dans lequel il est plongé. Or l'étude, c'est-à-dire l'identification, l'explication et la prévision de ce comportement sont précisément les objectifs de nombreuses démarches scientifiques et concernent aussi bien l'agent plus abstrait des sciences cognitives que celui, social, économique et biologique des sciences sociales et des sciences de la vie.

Comment préciser l'image de l'agent dans le cadre restreint que nous nous sommes imposé ? Tout d'abord, nous allons oublier l'image de la complexité réelle d'un acteur social, et ne retenir que le sujet pensant, l'agent cognitif. On peut alors assigner à cet agent idéal – comme le fait Johnson-Laird dans [4] – entre autres les tâches essentielles suivantes :

- percevoir le monde
- apprendre, se rappeler et commander des actions
- réfléchir et créer de nouvelles idées
- gérer la communication avec d'autres
- créer les sentiments, les intentions et la conscience de soi

Le problème de la perception est celui de l'acquisition d'information sur le monde et la représentation de celui-ci dans un modèle abstrait. Cela implique une codification des stimuli entrants et la réorganisation de ces informations en une structure imitant la réalité extérieure. La vision animale est l'exemple paradigmatique d'un tel processus. Chez un agent abstrait, il s'agira d'un simple système de senseurs ou de capteurs, réels ou virtuels, qui permet le passage de données caractéristiques, mais codées, de l'environnement vers le « cerveau » de l'agent. Ces données permettront d'élaborer une représentation interne du monde extérieur. Cette représentation à son tour permet à l'agent de s'adapter à son univers et de réagir à des dangers ou sollicitations extérieurs ou de prendre des initiatives. Elle lui donne les moyens d'apprendre.

Apprendre, c'est associer. Mais cela suppose la capacité de créer des liens entre objets situés à des endroits différents de l'espace et du temps et celle d'utili-

ser l'histoire passée pour améliorer les possibilités adaptatives de son comportement. L'apprentissage est donc fortement lié à la notion de *mémoire*. Celle-ci est analysée habituellement, un peu arbitrairement peut-être, en mémoire sensorielle immédiate, en mémoire de travail à court terme et en mémoire « permanente » ou mémoire à long terme. Cette faculté de retenir et comparer des événements est une condition nécessaire pour construire et exploiter une représentation. L'acquisition de nouveaux concepts par induction, l'enchaînement des idées et les diverses formes d'initiative par déduction, dépendent essentiellement de l'existence de cette mémoire.

La représentation schématique que l'on peut se faire de l'agent – de l'agent abstrait, idéal pour l'instant – comportera, de plus, les caractéristiques suivantes : ses actions sont engendrées par une structure hiérarchique de buts à atteindre et d'une base de connaissance lui permettant de produire des stratégies pour atteindre ces buts. Les agents eux-mêmes sont supposés être *rationnels* c'est-à-dire qu'ils utilisent une procédure déductive du type de celle décrite par une logique formelle quelconque pour établir leurs plans. Ces plans sont construits en confrontant les informations sensorielles aux divers buts envisagés à travers la représentation interne. L'organisation d'un agent est modulaire : chaque tâche énumérée plus haut correspond à un module séparé.

COMPORTEMENT

L'autonomie de l'agent et son caractère *situé*, qui l'intègre dans un environnement, le rendent a priori imprévisible pour un observateur. La plupart du temps, un certain degré de complexité ou de simplicité sera spontanément attribué à l'agent par l'observateur. Mais l'étude plus précise de son comportement sera d'un intérêt particulier et elle posera la question à la fois de la description et de l'explication de ce comportement. La description dépendra largement des facultés et modes de représentation de l'observateur. L'explication d'un comportement sera plus délicate et pourra se faire en trois ou quatre étapes.

L'explication à *court terme* essaiera d'élucider les causes immédiates d'un comportement local. Dans le cas de la fourmi il s'agira de comprendre une particularité de la courbe de sa trajectoire en fonction des accidents de l'environnement dans lequel celle-ci se déplace. Les règles suivies par la fourmi pourraient être : « contourner un obstacle par la droite s'il est situé devant et légèrement à gauche », « revenir en arrière en cas de cul-de-sac » etc. Ces règles pourraient constituer une explication à court terme en présence d'une description de l'état de la fourmi et de la configuration locale de l'environnement.

Dans une explication *ontogénétique* il entrera en plus un fragment de l'histoire de la fourmi. Des événements passés et mémorisés peuvent intervenir dans la recherche des causes d'une particularité du comportement. Ce type d'explication fait référence aussi à l'apprentissage. Notre fourmi aura appris par exemple que certains obstacles cachent ou annoncent d'autres difficultés et qu'il vaut mieux les éviter par un détour très large.

Dans une perspective *phylogénétique*, c'est l'histoire, non pas de la fourmi individuelle, mais de son espèce qui intervient. Certaines particularités du com-

portement s'expliquent par la structure même de la fourmi, plus généralement de l'agent, comme les caractéristiques des senseurs, ou son intelligence, voire son instinct. Ainsi certaines odeurs spécifiques vont l'attirer irrésistiblement ou alors la repoussent fortement.

On pourrait rajouter une quatrième forme, l'explication *téléologique* ou fonctionnelle. Elle répond aux questions : à quoi sert cette action ? vers quel but tend ce comportement ?

On peut rapprocher aisément ces quatre formes d'explication de la distinction faite par Aristote des quatre causes : matérielle, formelle, efficiente et finale.

Voici un exemple tiré de la vie quotidienne qui illustre les différents types d'explication d'un comportement⁵ : Observons un automobiliste qui s'arrête à un feu rouge. Une explication à court terme mettrait en avant la règle ou le réflexe suivant : feu rouge, donc pédale de frein et arrêt. Une explication ontogénétique insisterait sur l'apprentissage de l'automobiliste : celui-ci a appris à s'arrêter lorsque le feu est rouge, alors qu'une explication phylogénétique renverrait à l'histoire de la plupart de nos sociétés industrialisées et montrerait comment progressivement le feu rouge a été utilisé pour codifier le signal d'arrêt. L'explication téléologique par contre révélerait le simple désir d'éviter un accident ou un conflit avec la police.

Le phénomène d'émergence est particulièrement intéressant dans le cas d'un système social, c'est-à-dire dans le cas d'un système dont les parties sont eux-mêmes des agents⁶. Comment peut-on comprendre le comportement global à partir du comportement des agents individuels ? La fourmilière est un exemple typique de cette situation. Le plan et le fonctionnement de celle-ci n'est inscrit ni codé nulle part dans les agents et sa mise en place n'est qu'une émergence de l'interaction des différentes fourmis, présentes en grand nombre. Les phénomènes émergents sont, dans la majorité des cas, à l'œuvre dans des systèmes non linéaires à grande échelle.

Ces processus collectifs ont intrigué les biologistes depuis longtemps, particulièrement dans le cadre des études des sociétés animales, de la morphogenèse et de la théorie de l'évolution darwinienne. Et c'est avec un travail sur le développement de l'embryon [11] que le mathématicien anglais Alan Turing a jeté les bases d'une science mathématique de la complexité⁷. Celle-ci permet jusqu'à un certain point de décrire et d'expliquer formellement l'apparition et l'évolution de ces comportements globaux, surgissant de l'interaction d'un grand nombre d'agents entre eux et avec leur environnement. Cette théorie a été appliquée depuis non seulement à la biologie et à la physique, mais aussi à la psychologie sociale, à l'économie⁸, à la sociologie, à la géographie des développements urbains⁹ et à bien d'autres sciences sociales.

⁵ Martin & Bateson, cité dans Pfeiffer & Scheier [8]

⁶ Un système appelé multi-agents.

⁷ Voir pour cela les travaux de Shannon, von Neumann, Weaver, Wiener, Prigogine, Holland entre autres.

⁸ Voir Paul KRUGMAN, *The self-organizing economy*, Blackwell 1996

⁹ Voir le livre de Jane JACOBS, *The death and life of great american cities*, Vintage 1961 et le texte de Denise Pumain dans ce volume.

On a tenté, avec beaucoup d'audace, d'appliquer ces mêmes idées à l'esprit humain, le considérant lui aussi comme système social dans le sens décrit plus haut. Marvin Minsky [8] a présenté une vision cohérente de ce point de vue. Voici comment il décrit sa représentation de l'esprit :

J'appellerai « Société de l'esprit » ce système selon lequel chaque esprit est composé d'un grand nombre de petits processus que nous appellerons agents. Chaque agent ne peut, à lui seul, effectuer que quelques tâches simples ne demandant ni esprit ni réflexion. Pourtant, le regroupement de ces agents en sociétés – selon des modalités bien particulières – peut aboutir à la véritable intelligence.

Minsky donne alors une description détaillée d'une possible organisation de processus élémentaires, dénués d'intelligence, de laquelle pourrait émerger *l'esprit*. Même si le livre reste spéculatif, il décrit de manière plausible et détaillée comment un comportement humain complexe peut être interprété comme résultat d'une série d'interactions entre agents primitifs.

Mais il existe des agents plus proches encore de l'existence matérielle, qui illustrent parfaitement ce phénomène dans un contexte à la fois plus général et plus modeste. C'est notre exemple suivant.

VÉHICULES DE BRAITENBERG : SCIENCE ET FICTION

Dans un livre ludique [1], novateur et singulièrement stimulant, Valentino Braitenberg, spécialiste de neuro-sciences, propose une expérimentation, plus exactement un « Gedankenexperiment », avec des agents simples qu'il appelle des *véhicules*. Le but, légèrement subversif, du livre étant de montrer l'apparition spontanée de concepts, d'idées, de raisonnements, d'initiatives voire de sentiments et de conscience chez des agents artificiels rudimentaires, mais dont le comportement semble imiter de manière schématique celui des animaux et parfois même de l'homme. La grande simplicité de ces agents permet de les considérer comme des vues de l'esprit et ainsi de les manipuler mentalement, ou alors de les représenter sous forme de dessin, mais aussi, pourquoi pas, de les construire matériellement, donc de leur donner, en tant que robots, une existence incarnée, incorporée.

Braitenberg commence son livre par la phrase :

Ceci est un exercice d'imagination scientifique, ou de science-fiction si vous préférez

Mais il ne s'agit pas de science-fiction en tant que divertissement. Il s'agit d'une science-fiction au service de la science. Alors disons tout simplement de la science, si vous êtes d'accord avec le fait que la fiction en fait partie et en fera toujours partie; aussi longtemps que nos cerveaux ne seront que de minuscules fragments de l'univers, beaucoup trop petits pour contenir tous les faits du monde, mais suffisamment significatifs pour spéculer à leur sujet.

Les agents définis dans ce livre, les véhicules dans la terminologie de Braitenberg, sont caractérisés par une structure interne primitive, facile à comprendre, mais engendrent cependant des comportements variés et parfois très complexes.

Pour un observateur extérieur, ces comportements s'apparentent à quelques grands types de comportements qu'on retrouve chez les animaux et même les hommes et qui semblent motivés par l'agressivité, la peur, l'amour et l'existence de préférences, de valeurs et de goûts, mais aussi par la présence d'une forme élémentaire de raisonnement.

Les véhicules imaginés par Braitenberg possèdent des senseurs capables de percevoir des aspects physiques du monde qui les entoure. Ces capteurs sont reliés à des moteurs qui assurent leur mobilité et leur permettent de cette manière élémentaire d'exprimer leurs réactions face à l'environnement. Ces agents satisfont ainsi à certaines des exigences de base que Johnson-Laird impose à la notion d'agent, notamment les facultés de perception et d'action. Braitenberg présente quatorze véhicules dans un ordre croissant de sophistication. Le point important est que le progrès d'un véhicule à l'autre provient pour l'essentiel de la nature et de la qualité de connexions, plus que d'une amélioration des capteurs et moteurs eux-mêmes. Nous allons esquisser brièvement les premières étapes de cette progression.

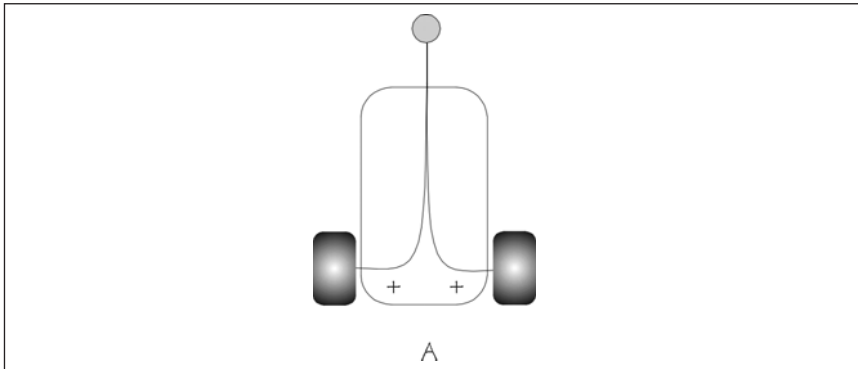


Figure 1.

Le plus sommaire des véhicules de Braitenberg est le véhicule A (figure 1). Supposons que son unique senseur soit sensible à la lumière. Le signe «+» signifie que le véhicule transforme une plus grande intensité lumineuse captée par un senseur en augmentation de la vitesse de son moteur. Il accélère donc en présence de lumière plus forte et ralentit dans les zones moins éclairées. Il semble par conséquent, vu de l'extérieur, que celui-ci apprécie la pénombre et cherche au contraire à échapper à la lumière vive. On lui prête aisément une *intention*. Le véhicule se déplace théoriquement en ligne droite mais on peut l'imaginer aussi dans un milieu qui n'est pas parfaitement homogène. Il pourrait en outre subir des frottements dus à de nombreuses forces minimes et aléatoires. Dans ce cas, la présence de ces forces parasites faibles peut faire apparaître la trajectoire comme très complexe. L'observateur constatera dans ce cas que le véhicule n'aime pas la lumière et qu'il recherche à la fuir, mais aussi qu'il suit une trajectoire incurvée, ou accidentée sans raison apparente, qu'il ne comprendra pas et qu'il pourra encore moins prédire.

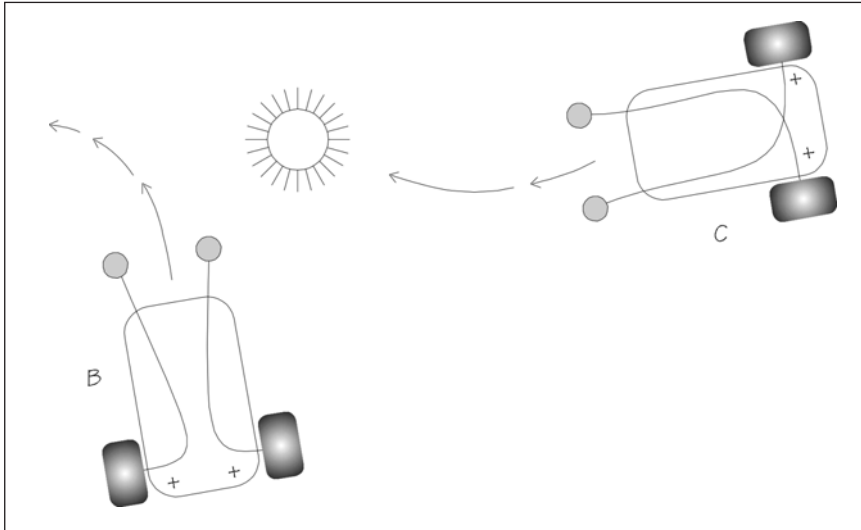


Figure 2.

Le véhicule B (figure 2) ne semble guère plus évolué, sauf qu'il est équipé de deux senseurs séparés. Cela introduit un grand progrès dans son expérience de la géométrie du monde puisque cela lui permet une perception de la direction. Comme les deux senseurs agissent séparément sur les deux moteurs, le véhicule est libre de changer le sens de son déplacement en fonction des événements. Que se passe-t-il par exemple lorsque le véhicule B s'approche d'une source lumineuse située devant lui et légèrement sur sa gauche? Le senseur de gauche, plus proche de la source lumineuse, va induire une accélération plus forte dans le moteur de gauche et le véhicule évitera la lumière d'abord très rapidement, puis de moins en moins vite. Pour un observateur, le véhicule semble posséder une personnalité craintive qui tend à éviter et fuir avec vivacité les sources lumineuses puis se calme progressivement dans la pénombre en s'éloignant.

Au contraire le véhicule C (figure 2) sera perçu comme particulièrement agressif! Pourtant la seule différence qui le distingue de B est la manière dont les senseurs sont reliés aux moteurs. Il s'agit ici d'une liaison croisée, un *chiasme* dans la terminologie anatomique, qui rappelle entre autres le croisement des nerfs optiques chez l'homme. Mais cette simple inversion des connexions va bouleverser le comportement du véhicule. En effet, lorsqu'il détecte une source lumineuse, le véhicule C se dirige droit sur celle-ci en accélérant. L'image parfaite d'une agression! Le véhicule déteste la lumière et cherche cette fois avec vigueur à détruire sa source. Telle est du moins l'interprétation sémantique plausible de la scène qui s'impose à un observateur extérieur.

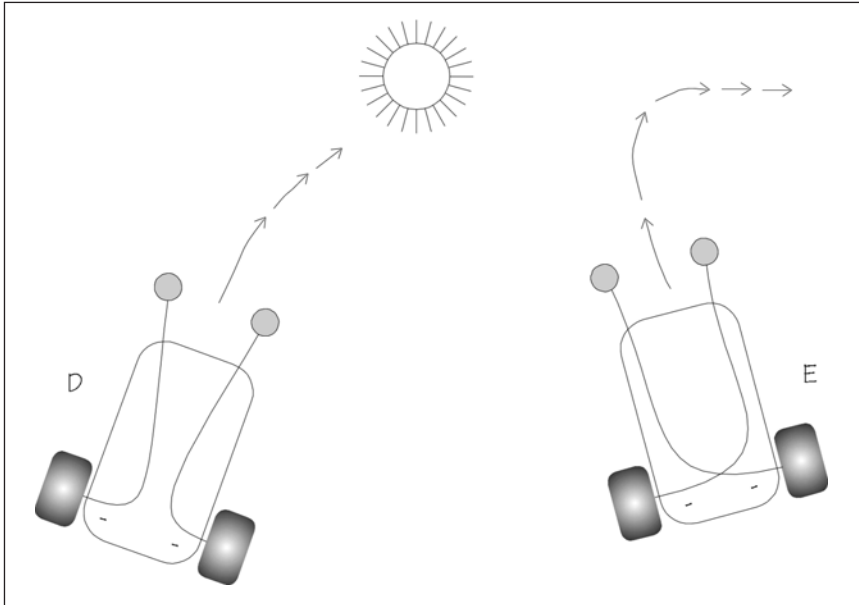


Figure 3.

Si les connexions ne sont pas excitatrices, mais au contraire inhibitrices, c'est-à-dire qu'elles ralentissent les moteurs avec un stimulus grandissant, alors le comportement est à nouveau complètement changé. Le véhicule D (figure 3) se précipite vers la lumière, puis ralentit et s'arrête en face, très près de la source lumineuse. Il semble « aimer » la lumière, au point de la rechercher fiévreusement et de se complaire dans le voisinage immédiat d'une lampe.

Un croisement des connexions dans le même cas produira chez le véhicule une autre forme de comportement. Le véhicule E (figure 3) s'approche rapidement d'une source, puis se détourne d'elle en ralentissant et en s'arrêtant éventuellement près d'elle, mais en lui tournant le dos, prêt à se diriger vers une autre source lumineuse plus lointaine qui pourrait se présenter dans son champ de vision. Les deux véhicules D et E « aiment » la lumière, mais pas de la même manière ! Braintenberg décrit cette différence à peu près dans ces termes :

D aime de façon permanente, en une admiration béate, alors que E est un explorateur, il aime la lumière et se complait dans sa proximité mais garde un œil ouvert pour d'autres sources !

Le véhicule F (figure 4) est un exemple de véhicule multisensoriel, il peut réagir aussi bien à la lumière qu'à la proximité d'un obstacle. Supposons qu'il soit équipé de senseurs sensibles à la lumière (ronds) et de senseurs de proximité, repérant les obstacles (carrés). Son comportement est plus diversifié que celui des véhicules précédents. Ainsi il évite les obstacles et se détourne de la lumière. Mais situé exactement en face d'une source lumineuse, il attaque celle-ci de front en

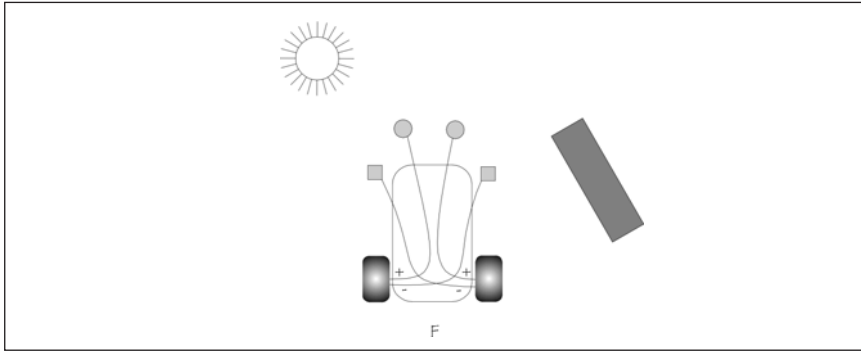


Figure 4

accélérant tout d’abord, puis en s’arrêtant à proximité de la source. D’autre part, malgré sa construction sommaire, il est capable de trouver son chemin pour sortir de certains types de labyrinthes.

Jusqu’ici nous n’avons que peu parlé des propriétés spécifiques de la connexion entre senseur et moteur, sauf pour distinguer deux fonctionnements différents : plus l’intensité est forte et plus la vitesse augmente, ou dans l’autre cas, plus l’intensité augmente, plus le véhicule ralentit. Ces connexions peuvent être modélisées par une fonction mathématique *monotone*. Si l’on admet que des fonctions non-monotones soient réalisées dans les connexions d’un véhicule, le comportement de ces véhicules peut alors paraître d’une richesse incroyable. Et c’est dans ce sens que Braitenberg poursuit. Voici quelques possibilités de développements nouveaux. Le véhicule peut par exemple être attiré par une source lumineuse et s’en approcher pour ralentir à proximité, puis s’en détourner pour la contourner le long d’un grand cercle. Il peut encore éviter cette source en décrivant un grand huit autour de lui. Le véhicule semble être habité par des *instincts*. Si la fonction de connexion possède des discontinuités, alors il peut arriver que le véhicule semble hésiter à prendre une décision. Il peut rester immobile un long temps avant de se précipiter soudain à grande vitesse sur une source lumineuse, puis de ralentir en parcourant une courbe complexe. *L’hésitation*, même la *réflexion* semblent donc faire partie de son fonctionnement. De même l’idée de *volonté* – suggérée uniquement par la forme de la fonction de transition – est présente à l’esprit de l’observateur qui tente d’interpréter le comportement d’un tel véhicule.

Mais comment franchir un nouveau pas et inculquer des « idées » à ces agents rudimentaires ? On peut, dans un premier temps, intercaler un réseau de dispositifs à seuil très primitif, une forme de neurones *artificiels*, entre senseur et moteur et créer ainsi une ébauche de cerveau en suivant la voie ouverte par McCulloch et Pitts dans [7]. Le véhicule, notre agent, devient alors capable de calculer¹⁰ et de réagir de manière « rationnelle » aux sollicitations de l’environnement. Il possède également une mémoire, qui, même rudimentaire, l’autorise à retenir de son

¹⁰ Techniquement cela signifie qu’il acquiert la capacité d’une machine de Turing

histoire passée les éléments importants qui lui permettent de mieux s'adapter au monde dans lequel il se trouve. Il est capable en plus capable *d'apprendre*.

En imaginant des modifications des propriétés de conduction¹¹ des connexions entre les éléments du «cerveau», Braitenberg montre comment les véhicules deviennent alors capables d'*association* et finalement de formation de concepts. Ils perçoivent également les *formes* et détectent le *mouvement*. La mémoire se développe et se spécialise. Elle permet de créer des cartes internes du monde extérieur telle qu'il est perçu par les capteurs du véhicule et utilisées pour calculer, puis effectuer des actions. Mais la mémoire rend aussi possible l'apparition d'un processus d'induction, ou du moins de généralisation. La possibilité de détecter des régularités donne aux véhicules la faculté de percevoir des *chaînes causales*. Les véhicules apprennent aussi à distinguer leurs «semblables» des autres parties de l'environnement.

Finalement, par des améliorations supplémentaires de la plasticité des connexions, Braitenberg parvient à créer des véhicules capables de *prévoir*. Il les soumet parallèlement à un processus de sélection naturelle qui lui permet de les «améliorer» sans utiliser de critère d'adaptation arbitraire. A la fin, les véhicules sont devenus des agents autonomes, situés, actifs, réactifs et sociaux. Les plus raffinées de ses créatures sont même capables d'un comportement qui suggère l'existence en eux de quelque chose qui ressemble à la *conscience*.

CONCLUSION

Même si l'on se tient à ce cadre limité aux sciences cognitives et à l'intelligence artificielle, les agents autonomes, tels que nous les avons rencontrés, présentent des comportements fascinants. Avant tout ils nous livrent cette leçon importante : la complexité ou l'intelligence qu'un observateur extérieur peut attribuer à ces comportements, n'est pas (uniquement) le reflet de la complexité de la structure interne des agents – ceux-ci peuvent être rudimentaires mais présenter un comportement très difficile à analyser et à prévoir. On peut, au contraire, considérer le comportement comme une *émergence* de l'interaction des agents avec l'environnement ou avec d'autres agents et avec l'observateur. Mais que faut-il entendre par là exactement ?

Le terme émergence est pris ici dans trois acceptions différentes, quoique liées :

- A. il désigne un phénomène qui n'est pas «préprogrammé» dans la structure cognitive de l'agent et qui se révèle spontanément au cours de la confrontation avec l'environnement ou avec d'autres agents
- B. il caractérise une propriété d'un système qui n'est présente dans aucune des parties de celui-ci, et qu'une analyse du système ne permet pas de prédire

¹¹ Par une astuce aussi amusante que convaincante, le fil Mnémotrix, caractérisé par une forme de renforcement introduisant la plasticité qui permet d'intégrer la loi d'apprentissage de Hebb dans le monde des véhicules.

C. il s'applique à des manifestations créant un effet de surprise et qui n'entrent pas dans un schéma explicatif connu

Cette notion d'émergence se retrouve aussi bien chez Platon lorsque il distingue, dans le *Théétète*, le tout comme somme des parties et le tout comme émergence des parties. Elle se retrouve aussi, par exemple, chez Mill dans le problème de la composition des causes. Mill distingue le cas où l'effet produit peut-être considéré comme la somme des effets – qu'on peut appeler effet *résultant*, de celui où l'effet est de nature entièrement différente des effets séparés – qui serait l'effet *émergent*.

Dans la philosophie des sciences, c'est le plus souvent la définition B qui a été prise en compte. Le caractère non prédictible d'un phénomène émergent a été largement utilisé comme argument contre le (micro-) réductionnisme, qui prétend ramener toute l'étude du comportement global d'un système à celui de ses parties. Mais cet argument perd sa plausibilité et sa force si l'on distingue clairement *explication* et *prédiction*. Car si effectivement un phénomène émergent n'est pas prévisible, il peut tout de même, dans un grand nombre de cas et jusqu'à un certain degré, être expliqué. Le chemin de la fourmi sur le sable peut être mieux compris si une micro-analyse – éventuellement laborieuse – nous fournit tous les détails des obstacles que la fourmi rencontre à son échelle et qu'elle confronte ces données avec les règles de conduite internes de la fourmi.

La notion de phénomène émergent perd, par cette distinction explication/prévision, le caractère un peu mystique que pouvait lui valoir l'idée qu'un tel phénomène était mystérieux et hors de portée de notre compréhension. Même si l'on considère le comportement d'un agent comme émergent, celui-ci n'échappe pas entièrement à un observateur. Bien sûr, son comportement peut rester tout à fait imprédictible. Mais il se peut que par une analyse très fine, très détaillée, on trouve une suite d'explications à partir d'éléments plus petits du système formé par l'agent et par l'environnement dans lequel il est plongé.

Ainsi dans le monde économique, les phénomènes d'inflation ou de chômage, sont – du moins selon certaines théories économiques – des émergences, mais cela n'exclut pas la possibilité que par une micro-analyse très précise, on puisse isoler leurs causes et amorcer ainsi une explication de ces manifestations. La prévision, par contre, reste généralement en dehors du domaine du possible.

En biologie de l'évolution, certains phénomènes de type émergents semblent difficilement compatibles avec l'idée de sélection naturelle. C'est le cas notamment de l'apparition de l'œil chez les mammifères ou des ailes chez les oiseaux. Si la sélection naturelle est réellement « myope », comment expliquer les avantages évolutifs des petits pas qui ont conduit finalement à ces progrès qui constituent un saut qualitatif important ? Ici encore, une analyse de détail de l'interaction espèce/environnement, a permis dans une large mesure¹² de donner une explication conforme aux lois darwiniennes de la sélection naturelle.

Nos quelques remarques sur les notions d'agent autonome et de comportement ont porté le plus souvent sur un agent simplifié, parfois virtuel. Il n'en reste pas

¹² Voir les livres de Dawkins [2] et Dennet [3].

moins que la conception du comportement comme phénomène émergent, issu de la confrontation agent/environnement/observateur, peut se rapporter également aux agents humains¹³, tels qu'ils apparaissent dans notre contexte social. Dans ce cas, les choses ne seront plus aussi simples, mais la présence de quelques grands types de phénomènes émergents, comme l'auto-organisation, les transitions de phase ou la rupture de symétrie, pourront faire figure d'indicateurs précieux. Alors, bien que non-prévisible, le comportement des agents pourra faire l'objet d'une série d'explications obtenues par une analyse empirique minutieuse.

Les véhicules de Braitenberg, en dépit de leur insolente simplicité, ouvrent un magnifique champ de réflexion sur les conditions minimales dans lesquelles apparaissent ces phénomènes émergents qui correspondent à nos concepts explicatifs les plus importants parmi lesquels figurent : amour, agressivité, soif de pouvoir, défense de son territoire, regroupement en clan, méfiance, peur et bien d'autres. Et bien qu'ils soient d'un autre monde – essentiellement plus simple – que notre réalité sociale, ils pourraient bien constituer des indicateurs originaux, permettant d'orienter certaines recherches d'explication pour des comportements particulièrement intrigants. Tout comme la société de l'esprit de Minsky, malgré son apparence polémique et simplificatrice, peut induire une réflexion renouvelée sur le problème de l'esprit, basée sur l'idée d'émergence à partir d'une société d'agents aux possibilités très restreintes.

Alors agent ou acteur¹⁴? Dans une société concrète, complexe, c'est peut-être le phénomène d'émergence qui éclaire le mieux cette distinction, comme nous l'avons vu : à la fois individu autonome prenant des initiatives *locales*, et catalyseur collectif de formes de comportement beaucoup plus *globales*, il est acteur aussi bien qu'agent, en fonction du point de vue de l'observateur.

Ces quelques remarques sur la notion d'agent auront au moins permis, nous l'espérons, d'éclairer quelques points de définition. Si le point de vue des sciences cognitives a été généralement privilégié et si les exemples présentés ont été, par nature, relativement formels ou abstraits – donc simples, l'acteur social réel a néanmoins toujours été présent en filigrane. Peut-être que la distance engendrée par ce parti pris, pourra être interprétée comme recul salutaire, un peu ludique et stimulant dans le cadre théorique bien plus complexe des sciences sociales.

*Institut de mathématiques appliquées
Université de Lausanne*

BIBLIOGRAPHIE

- [1] Valentino BRAITENBERG, *Vehicles, Experiments in synthetic psychology*, MIT Press 1984. Trad. Française, *Véhicules, expériences en psychologie synthétique*, PPUR 1991.
- [2] Richard DAWKINS, *Climbing Mount Improbable*, Viking 1996.
- [3] Daniel C. DENNET, *Darwin's dangerous idea*, Simon & Schuster 1995.

¹³ Ceux-ci peuvent être des personnes, mais aussi des groupes de personnes, voire des groupes de groupes.

¹⁴ Voir la distinction faite dans la contribution de Jean-Pierre Gaudin.

- [4] Philip N. JOHNSON-LAIRD, *The computer and the mind*, Fontana Press 1993.
Trad. Française, *L'ordinateur et l'esprit*, Odile Jacob 1994.
- [5] Christopher G. LANGTON ed., *Artificial life. An overview*, MIT Press 1995.
- [6] P. MARTIN & P. BATESON, *Measuring behavior : An introductory guide*, CUP 1993.
- [7] W.S. McCULLOCH & W.H. PITTS, A logical calculus of ideas immanent in nervous activity, *Bull. Math. Biophys.* 5, 1943.
- [8] Marvin MINSKY, *The society of mind*, Simon & Schuster 1986.
Trad. Française, *La société de l'esprit*, InterEditions 1988.
- [9] Rolf PFEIFER & Christian SCHEIER, *Understanding intelligence*, MIT Press 1999.
- [10] Herbert A. SIMON, *The sciences of the artificial*, MIT Press, Cambridge 1982.
Trad. française, *Sciences des systèmes, sciences de l'artificiel*, Dunod 1991.
- [11] P.T. SAUNDERS ed., *Collected works of A.M. Turing, Morphogenesis*, North-Holland 1992.

(Remerciements à Marek Blaszczyk pour les illustrations)