

## REPRÉSENTATION ERRONÉE

L'épistémologie s'intéresse à la connaissance : comment arrivons-nous à saisir les choses correctement? Une autre question va plus loin : comment arrivons-nous à saisir les choses incorrectement? Comment est-il possible pour des systèmes physiques de mal représenter l'état de leur environnement?

Le problème n'est pas de savoir comment, par exemple, un diagramme,  $d$ , peut mal représenter le monde,  $m$ . Car si nous disposons d'un autre système,  $r$ , possédant déjà des pouvoirs représentationnels,  $d$  peut alors être utilisé par  $r$  à titre d'instrument ou d'outil expressif, et prendre part par le fait même aux succès et aux échecs représentationnels de  $r$ . Le cas échéant,  $d$  peut en venir à signifier que  $m$  est  $F$  bien qu'en fait  $m$  ne soit pas  $F$ , ce qu'il signifie, en définitive, découlant de  $r$ . Un schéma traçant la courbe du chômage au cours des dix dernières années peut faussement représenter cet état de chose, mais la capacité que possède le schéma de mal représenter un état du monde découle du rôle qu'il joue en tant qu'instrument d'expression pour des agents, des locuteurs d'un langage, qui possèdent déjà ce pouvoir.

Non, le problème concerne plutôt les pouvoirs représentationnels d'un système dans la mesure où ces pouvoirs ne dérivent pas des efforts de représentation provenant d'une autre source. À moins de posséder un indice quelconque révélant comment une telle situation est possible, nous n'avons aucune idée de la façon dont les systèmes biologiques qui ont évolué de manière naturelle aient pu acquérir la capacité de former de telles croyances. Car la croyance, du moins je le supposerai, est une capacité représentationnelle non dérivée dont la mise en œuvre peut donner lieu à une mauvaise représentation.

La capacité à mal représenter n'est qu'une partie, peut-être seulement infime, du problème plus général de la signification ou de l'intentionnalité. Une fois la signification à notre disposition, nous sommes à même, dans nos descriptions et nos explications du comportement humain et animal, et peut-être même celui de machines, d'en faire largement usage à l'égard des systèmes que nous décrivons. Une fois l'intentionnalité à notre disposition, nous sommes à même (pour emprunter le langage de Dennett) d'adopter la pose intentionnelle<sup>1</sup>. Mais qu'est-ce qui (à part l'intentionnalité) nous donne (à nous et non pas, disons, aux machines) le pouvoir d'adopter cette pose? Notre capacité à adopter la pose intentionnelle est une *manifestation* de l'intentionnalité, non son analyse. La signification d'emprunt de systèmes envers lesquels nous adoptons des attitudes appropriées ne nous en dit pas plus sur la capacité initiale de mal représenter qu'une épingle placée sur une carte militaire. Ce que nous recherchons, pour ainsi dire, c'est la manière dont la *nature* elle-même se trompe, ce à quoi la mauvaise représentation est finalement imputable. Ce n'est qu'à compter du moment où nous comprendrons cela que nous serons à même de comprendre comment la matière grise peut mal représenter le temps qu'il fera demain pour le pique-nique.

1. Daniel C. Dennett (1978). «Intentional Systems», in *Brainstorms*, Cambridge, MIT Press.

## 1. SIGNES NATURELS

Les signes advenant de façon naturelle font sens, et ce sans le moindre concours de notre part<sup>1</sup>. L'eau ne coule pas en amont; donc, une rivière coulant vers le nord signifie qu'il y a une inclinaison vers le bas dans cette direction. Des ombres projetées vers l'est signifient que le soleil est à l'ouest. Une force subite exercée dans une certaine direction sur les passagers signifie que le train accélère en direction inverse. L'aptitude que possèdent ces événements ou ces conditions à signifier ce qu'elles signifient est indépendante de la manière dont nous les interprétons – ou même de ce que nous les interprétons ou les reconnaissons. Une dentiste peut avoir recours à une radiographie afin de diagnostiquer l'état de votre molaire supérieure droite, mais les traces ombragées s'y trouvant signifient qu'un état de carie avancé est survenu – que la dentiste, ou quiconque, soit ou non à même de l'apprécier. Un métal en expansion indique un accroissement de la température (et à ce titre signifie que la température augmente), qu'on en vienne ou non à le croire en l'observant. C'est ce que cela signifiait avant que des organismes intelligents, capables d'exploiter ce fait (en fabriquant des thermomètres), n'aient habité la Terre. Si nous sommes à la recherche de la source ultime de la signification, si nous voulons comprendre le pouvoir de mauvaise représentation que possède un système, il semble bien que nous ayons ici un point de départ prometteur.

Les signes naturels sont des indicateurs, plus ou moins fiables, et ce qu'ils signifient est ce qu'ils indiquent être le cas. Le pouvoir que possède un signe naturel de signifier quelque chose

1. Cette assertion devrait être nuancée, mais devrait suffire pour le moment. Ce qu'un signe naturel signifie dépend souvent de nous, de ce que nous savons des possibilités de rechange pertinentes dont nous avons connaissance, ou de la manière dont nous *employons* un dispositif correspondant. Mais si nous ne savons rien, ou si le signe intervient dans l'opération d'un dispositif n'ayant aucun usage normal, le signe n'en signifie pas moins quelque chose – mais tout simplement pas la signification spécifique que nous lui attribuons dans des conditions épistémiques (ou fonctionnelles) plus riches. Je reviendrai sur ce point plus bas, à la note 4 page 95.

– que Julie a la rougeole, par exemple – est étayé par certaines contraintes objectives, certaines relations nomologiques, entre le signe (ou le fait que le signe possède une certaine propriété) et l'état qui constitue sa signification (le fait que Julie ait la rougeole). Dans la plupart des cas, cette relation est causale ou nomologique, et est à même de résister à un énoncé contrefactuel suivant lequel si l'un des états ne s'était pas réalisé (si Julie n'avait pas eu la rougeole), l'autre ne se serait pas réalisé non plus (elle n'aurait pas eu toutes ces taches rouges au visage). Parfois il n'y a que de simples régularités, prépondérantes bien que non nomologiques, qui contribuent à établir la connexion entre le signe et ce qu'il signifie. C'est en partie le fait, vraisemblablement de soi non nomologique, que les animaux (les écureuils et les pic épeiches, par exemple) ne sonnent pas régulièrement aux portes alors qu'ils cherchent de quoi se nourrir qui fait en sorte qu'une sonnette qui retentit *signifie* que quelqu'un (c'est-à-dire une *personne* quelconque) est à la porte. Si les écureuils modifiaient leurs habitudes (parce que les sonnettes étaient faites à base de noix, disons), alors une sonnette qui retentit ne signifierait plus ce qu'elle signifie maintenant. Mais la situation est à *présent* telle que nous pouvons (d'ordinaire) dire que la sonnette ne retentirait pas à moins que quelqu'un ne soit à la porte, que la sonnette indique la présence de quelqu'un à la porte, et que, par conséquent, c'est ce que le retentissement de la sonnette signifie. Mais le fait que cette dépendance entre le retentissement de la sonnette et la présence de quelqu'un à la porte soit exprimée au conditionnel reflète une régularité qui, bien que non conventionnelle, n'est pas complètement nomologique non plus. La sonnette conserve toutefois sa signification naturelle pour peu que cette régularité ait cours.

Au delà de ce qui précède, je n'ai rien de très systématique à dire à propos de ce qui constitue la signification naturelle d'un événement ou d'un état<sup>1</sup>. Je vais poursuivre avec une notion qui sera, je l'espère, raisonnablement familière, m'appuyant

1. J'en donne une description plus complète aux chapitres 1 et 2 de mon livre *Knowledge and the Flow of Information*, Cambridge, MIT Press, 1981.

(lorsque nécessaire) sur des exemples concrets. Le projet est de voir jusqu'où on peut aller afin de comprendre la mauvaise représentation – le pouvoir qu'a une condition (un état, un événement, une situation) *r* de *faussement* signifier (indiquer, disons) que *m* est *F* (et par le fait même représenter *m* de manière erronée) – en termes d'un signe naturel signifiant que *m* est *F*. (Ce ne sera qu'à partir du moment où le projet atteindra son but, ou à tout le moins en montrera des signes raisonnablement prometteurs, qu'il deviendra nécessaire de considérer plus attentivement ce qui a été introduit subrepticement dès le départ.

Bien que la signification naturelle soit un point de départ prometteur, il est difficile de voir comment mettre les choses en branle. Les signes naturels, bien qu'ils signifient quelque chose et qu'ils soient à même (en ce sens) de représenter *m* (en indiquant, ou en signifiant, que *m* est *F*), sont incapables de représenter *de manière erronée* quoi que ce soit. De deux choses l'une : ou ils réalisent leur tâche correctement, ou ils ne la réalisent pas du tout. Les taches au visage de Julie peuvent certainement signifier qu'elle a la rougeole, mais elles ne le signifient *que* lorsqu'elle a la rougeole. Si elle n'a pas la rougeole, les taches n'ont pas cette signification. Le cas échéant, tout ce qu'elles signifient, peut-être, c'est que Julie a mangé trop de sucreries.

Grice formule ce point en indiquant qu'une occurrence (l'occurrence d'un signe naturel quelconque) signifie (au sens naturel du terme, comme il l'appelle – ci-après signification<sub>n</sub>) que *P* seulement si *P*<sup>1</sup>. Il distingue ce type de signification de la signification non naturelle, où un signe peut signifier que *P* même si *P* est faux. Si nous réservons le terme « signification » (sans indice) pour les cas où quelque chose peut signifier que *m* est *F* lorsque *m* n'est pas *F* – le type de signification qui rend possible la représentation erronée – alors la signification<sub>n</sub> apparaît comme une bien piètre candidate pour qui cherche à comprendre ce qu'est la signification.

1. Paul Grice (1957), « Meaning », *Philosophical Review*, 66, p. 377-388.

Dans la mesure où je parle de signes ainsi que de leur signification naturelle, on devrait toujours entendre que je fais référence à des événements, des conditions ou des états particuliers : *cette trace-ci, ces nuages-là, cette fumée-là*. Un signe type (la fumée, par exemple) peut bien signifier, en un sens naturel, qu'il y a du feu même lorsque chacune des occurrences de ce type manque à le signifier<sub>n</sub> (parce qu'à l'occasion, il n'y a pas de feu). Mais peu importe son analyse précise, un tel cas de signification associée à un type ne peut nous aider à comprendre la fausse représentation que si les occurrences individuelles de ce type possèdent bien la signification associée à ce type – que si les bouffées de fumée particulières signifient<sub>n</sub> qu'il y a du feu alors qu'il n'y en a pas. Cela n'est cependant pas le cas. Une jauge d'essence qui marque « vide » (ce type d'événements) peut signifier que le réservoir est vide, mais lorsqu'il ne l'est pas, aucune instance particulière de la jauge marquant « vide » ne signifie<sub>n</sub> que le réservoir est vide. Par conséquent, aucune marque particulière de la jauge ne représente faussement la quantité d'essence dans le réservoir (en signifiant<sub>n</sub> qu'il est vide alors qu'il ne l'est pas).

L'incapacité des signes naturels (particuliers) à représenter de manière erronée quoi que ce soit est parfois masquée par la manière dont nous en tirons parti à l'aide de dispositifs manufacturés. Curieux de savoir si *m* devient *F*, et le cas échéant à quel moment, nous fabriquons un dispositif *d*, dont les états variés sont conçus pour fonctionner à titre de signes naturels de l'état de *m*. Puisque nous utilisons le dispositif à ces fins, nous avons tendance à déclarer, en présence d'une lecture particulière du marqueur indiquant que *d* est *G* (supposant qu'il s'agisse là du signe naturel pour les cas où *m* est *F*), que cela signifie que *m* est *F* même lorsque le système, en raison d'une défaillance ou d'un usage inadéquat, manque à accomplir sa tâche convenablement et que *m* n'est pas *F*. Il est clair, cependant, que ce n'est pas ce que la position particulière du marqueur sur le dispositif signifie<sub>n</sub>. C'est ce qu'elle est censée signifier<sub>n</sub>, ce qu'elle a été

conçue pour signifier<sub>n</sub>, ce que (peut-être) des occurrences d'un type signifient<sub>n</sub>, mais ce n'est pas ce qu'elle signifie<sub>n</sub> en fait.

Lorsque survient un court-circuit, le retentissement d'une sonnette (indépendamment de ce qu'il a été conçu pour indiquer, ou même de ce qu'il indique normalement) n'indique pas que le bouton de sonnette est pressé. Le retentissement signifie<sub>n</sub> (indique) toujours qu'il y a du courant électrique qui se déplace dans le circuit de la sonnette (ce qu'entre autre il signifie<sub>n</sub> en permanence), mais ce dernier ne signifie<sub>n</sub> plus que le bouton de sonnette est pressé. Ce que le déplacement du courant signifie<sub>n</sub> à présent – et c'est sans l'ombre d'un doute un jugement que nous porterions si nous étions à même de voir que le bouton de sonnette n'est pas pressé – est que le système est défaillant, ou qu'un court-circuit est survenu quelque part dans l'installation électrique. L'énoncé « Il y a quelqu'un à la porte » peut signifier qu'il y a quelqu'un à la porte bien que personne ne s'y trouve, mais le retentissement de la sonnette ne peut le signifier alors que personne n'y est. Du moins, pas lorsque nous considérons la signification<sub>n</sub>. Si le bouton de sonnette n'est pas pressé, nous devons chercher ailleurs afin de déceler ce que le retentissement de la sonnette signifie<sub>n</sub>. Souvent, nous nous rabattons sur une signification<sub>n</sub> proximale, soit un état de chose se trouvant au sein de la chaîne des antécédents causals habituels qui est en fait réalisé (le déplacement du courant ou ce qui le cause – un court-circuit, par exemple), et nous le désignons comme la signification<sub>n</sub> de la sonnerie.

## 2. SIGNIFICATION FONCTIONNELLE

Quelqu'un pourrait toujours admettre que le retentissement de la sonnette ne peut signifier<sub>n</sub> que quelqu'un est à la porte alors que personne ne s'y trouve, mais néanmoins soutenir qu'en un sens apparenté de signification, la sonnerie signifie cela même s'il n'y a personne. S'il ne s'agit pas de signification naturelle (signification<sub>n</sub>), il s'agit certainement d'un parent proche.

Qu'il s'agisse ou non d'une acception apparentée, il s'agit certainement un type de signification lié à des systèmes, ou à des composantes de systèmes, qui ont des *fonctions* identifiables. Examinons à nouveau la jauge d'essence. Elle a une fonction d'acheminer de l'information sur la quantité d'essence. Lorsque tout se déroule bien, la position de l'aiguille est un signe naturel du contenu du réservoir. Lorsqu'elle pointe vers la gauche, cela signifie<sub>n</sub> que le réservoir est vide. Lorsqu'elle pointe vers la droite, cela signifie<sub>n</sub> que le réservoir est plein. Et ainsi de suite pour les positions intermédiaires. Mais parfois, rien ne va plus : les connexions se desserrent, l'accumulateur tombe à plat, les fils se brisent. La jauge se met à marquer « vide » alors que le réservoir est encore plein. Lorsque cela se produit, nous avons une tendance à dire que la jauge représente de manière erronée [*misrepresents*] le contenu du réservoir. La jauge dit que le réservoir est vide alors qu'il ne l'est pas. Elle *signifie* (parce qu'elle signifie<sub>n</sub>, bien entendu, mais elle signifie tout de même, en un certain sens) que le réservoir est vide.

Lorsque le fait que *d* est *G* [*d's being G*] est normalement un signe naturel de ce que *m* est *F* [*m's being F*], lorsque normalement c'est ce qu'il signifie<sub>n</sub>, on dira alors qu'il le signifie – que *m* soit ou non *F* – si *d* a pour fonction d'indiquer l'état de *m*. Appelons ce type de signification signification<sub>f</sub> – l'indice indiquant qu'il s'agit de signification dérivée fonctionnellement.

(S<sub>f</sub>) le fait que *d* est *G* [*d's being G*] signifie<sub>f</sub> que *m* est *F* = *d* a pour fonction d'indiquer l'état de *m*, et la manière dont il s'acquitte de cette fonction est, en partie, en indiquant que *m* est *F* par le fait qu'il (*d*) est *G*.

La position de l'aiguille sur la jauge défaillante signifie<sub>f</sub> que le réservoir est vide parce que la jauge a pour fonction d'indiquer la quantité d'essence dans le réservoir, et la manière dont elle s'acquitte de cette fonction est, en partie, en indiquant un réservoir vide lorsqu'elle marque « vide »<sup>1</sup>. Au même titre, et pour

1. Il sera évident, je l'espère, que mon propos ici ne porte pas sur le mot « vide » (ou sur la lettre « V ») qui pourrait apparaître sur la jauge. Ce symbole signifie

comme raison, le retentissement de la sonnette exprime (c'est-à-dire signifie<sub>f</sub>) que quelqu'un est à la porte même lorsque personne ne s'y trouve.

La réponse à la question de savoir si (S<sub>f</sub>) marque ou non un signe en avant relativement à notre démarche visant à naturaliser la signification (et ainsi comprendre le pouvoir non-dérivé de la représentation erronée que possède un système) dépend de ce que les dites fonctions sont oui ou non elles-mêmes susceptibles d'être entendues, ou saisies, de façon naturelle. Si elles sont (ce que je nommerai) des fonctions *assignées*, alors la signification<sub>f</sub> sera entachée des buts, des intentions et des croyances de ceux et celles qui assignent la fonction d'où la signification<sub>f</sub> dérive sa capacité de fausse représentation<sup>1</sup>. Ce faisant, nous n'aurons pas réussi, dans la mesure où cela implique la capacité de représenter de manière erronée, à suivre la trace de la signification jusqu'à son point d'origine. Nous serons tout simplement revenus, quoique par un chemin quelque peu oblique, à l'énigme de *notre propre* capacité de représenter.

Pour mieux saisir ce que j'entends par fonction *assignée*, de même que la manière dont nous sommes impliqués (nos intentions, nos buts et nos croyances) lorsqu'un système possède une telle fonction, examinons le cas suivant. Imaginons une balance à ressort extrêmement sensible, calibrée au centième de grammes près, conçue et utilisée afin de déterminer le poids de menus objets. À l'insu des concepteurs et des utilisateurs, l'instrument indique l'altitude, et ce de manière tout aussi sensible. Comme son marqueur indique un poids réduit pour les objets au fur et à mesure que l'altitude augmente (N.B. : le poids d'un objet est fonction de son altitude au-dessus du niveau de la mer), l'instrument pourrait être utilisé à titre d'altimètre rudimentaire si l'utilisateur y joignait un poids standard et notait ses variations selon le changement d'altitude. Supposons maintenant que dans

« vide » peu importe comment la jauge se comporte, mais cela est purement conventionnel. Mon propos porte sur ce que la position de l'aiguille signifie<sub>n</sub> peu importe ce que nous choisissons d'afficher sur l'instrument.

1. Dans son article « Functions » (*Philosophical Review*, 82, 1973, p. 139-168), L. Wright les appelle fonctions « conscientes ».

des conditions normales en laboratoire, l'instrument défaille et marque 0,98 g pour un objet pesant 1 g. L'instrument représente-t-il de manière erronée le *poids* de l'objet? Représente-t-il faussement son *altitude*? Que signifie la lecture de 0,98 g? Si nous considérons la signification<sub>n</sub>, il est clair que la lecture ne signifie<sub>n</sub> pas que l'objet pèse 0,98 g. Et elle ne signifie<sub>n</sub> pas non plus que le laboratoire est situé à 14000 mètres au-dessus du niveau de la mer. Si cependant nous considérons la signification<sub>f</sub>, il semble alors raisonnable de dire que le marqueur de l'instrument exprime, ou indique (c'est-à-dire signifie<sub>f</sub>) que l'objet pèse 0,98 g. Cet instrument a pour fonction de nous signaler le poids d'un objet, et il nous signale (incorrectement, en l'occurrence) que cet objet pèse 0,98 g.

Mais l'altitude est-elle faussement représentée? Non. Il faut noter que l'instrument ne peut faussement représenter à la fois l'altitude et le poids, puisqu'une représentation (ou une représentation *erronée*) de l'un présuppose l'invariance (et donc, la *non-représentation*) de l'autre<sup>1</sup>. Bien que l'instrument pourrait servir d'altimètre, on ne s'en sert *pas* ainsi. Il ne s'agit pas là de sa fonction. Sa fonction consiste à marquer le poids. C'est là la fonction que nous lui assignons; la raison pour laquelle il a été fabriqué; ce qui explique pourquoi il a été fabriqué de cette façon et non d'une autre. Aurions-nous eu d'autres buts, il eût pu signifier<sub>f</sub> quelque chose d'autre. Mais tel n'est pas le cas.

De temps à autre nous modifions la fonction assignée d'un instrument. Lorsque nous le calibrons, par exemple, nous cessons de l'utiliser afin de mesurer ce qu'il mesure normalement. Nous le confrontons plutôt à des quantités connues afin d'utiliser ce qu'il indique comme signe (naturel) marquant sa possible défaillance ou imprécision. En ce cas, une lecture de 0,98 g (pour un poids que l'on sait être de 1 g) indique que les caractéristiques du ressort ont changé, que l'aiguille a fléchi, ou qu'une autre composante est dérégulée. Nous nous retrouvons avec une nouvelle signification fonctionnelle parce que notre

1. Une sonnette, par exemple, ne peut signifier<sub>n</sub> à la fois qu'il y a quelqu'un à la porte et qu'il y a un court-circuit.

connaissance d'arrière-plan, maintenant altérée (qui résulte normalement d'intentions et de buts différents), a pour effet de modifier ce que le comportement du marqueur signifie<sub>n</sub>. Avec les fonctions *assignées*, les significations<sub>f</sub> changent au fur et à mesure que nos buts changent<sup>1</sup>.

Nous utilisons parfois des animaux de la même façon que nous utilisons les instruments. Les chiens ont l'odorat fin. Les douaniers tirent parti de ce fait en utilisant des chiens afin de déceler de la marijuana. Lorsqu'un chien agite la queue, aboie ou accomplit ce qu'il a été dressé pour accomplir lorsqu'il flaire de la marijuana, son comportement sert de signe naturel – un signe à l'effet que les bagages contiennent de la marijuana. Mais cela ne signifie pas que le comportement du chien (ou l'état neuronal qui déclenche ce comportement) peut représenter de manière erronée le contenu. Le comportement du chien peut conduire la douanière à croire (à tort) qu'il y a de la marijuana dissimulée dans les bagages, mais ce comportement ne le signifie<sub>f</sub> que de manière dérivée. Si le chien est tout particulièrement doué, n'aboyant qu'en présence de marijuana, nous pouvons dire que son aboiement indique (c'est-à-dire signifie<sub>n</sub>) la présence de marijuana. En outre, c'est ce que l'aboiement signifie<sub>n</sub> – qu'on l'interprète comme tel ou non; que nous utilisions ou non ce signe naturel pour nos propres fins. Mais lorsqu'il n'y a pas de marijuana, lorsque le chien aboie en présence d'une anodine caisse d'aromates, l'aboiement ne signifie<sub>n</sub> *pas* la présence de

1. Ce n'est pas la modification du but ou de l'objectif seule qui modifie ce que quelque chose signifie<sub>n</sub> (et donc signifie<sub>f</sub>). C'est le fait que ce changement d'usage s'accompagne d'une connaissance d'arrière-plan altérée, et que la signification<sub>n</sub> change au fur et à mesure que la connaissance d'arrière-plan change. Si, par exemple, A dépend à la fois de B et de C, un changement en A peut signifier<sub>n</sub> que C a changé si nous savons que B est demeuré constant. Si nous savons que C est constant, cela peut signifier<sub>n</sub> que B a changé. Mais si nous ne possédons pas cette connaissance, cela signifie seulement que soit B, soit C a changé. La signification naturelle est en ce sens relative, mais la dérelativiser (en ne tenant pas compte de ce dont nous avons connaissance et comment nous utilisons un dispositif) ne la fera pas disparaître pour autant. Cela n'aura tout simplement pour effet que ce que les choses signifient<sub>n</sub> sera moins déterminé. Pour une discussion plus soutenue sur ce point, voir le chapitre 3 de Dretske (1981), *op. cit.*

marijuana. Il ne le signifie<sub>f</sub> pas non plus en un quelconque sens qui serait indépendant de nos efforts interprétatifs. Nous pouvons bien entendu exprimer ce que l'aboiement signifie *pour nous* (qu'il y a de la marijuana dissimulée dans la valise), mais cette façon de parler révèle tout simplement le rôle que nous jouons dans l'attribution de la signification du comportement du chien. Nous lui attribuons cette signification parce qu'il s'agit de l'information qui nous *intéresse*, l'information que nous *prévoyons* obtenir en utilisant le chien de cette façon, l'information que le chien a été dressé à transmettre. Mais si nous laissons de côté nos buts et nos intérêts, il n'y a *pas* de sens où l'aboiement du chien signifie qu'il y a de la marijuana dissimulée dans les bagages alors *qu'il n'y en a pas*. Le seul genre de représentation erronée advenant en ce cas est du type dérivé auquel les cartes, les instruments et le langage nous ont familiarisés.

Il s'ensuit que si  $S_f$  doit servir d'explication naturaliste de la représentation, y compris les cas de représentations *erronées*, les fonctions en question doivent être des fonctions naturelles, des fonctions qui sont indépendantes de *nos* intentions et de *nos* buts interprétatifs. Ce que nous recherchons, c'est des fonctions qui impliquent un système de signes naturels et qui confèrent un contenu à ces signes, et donc une signification (c'est-à-dire une signification<sub>f</sub>) qui n'est parasitaire ni de la manière dont nous les exploitons lors d'activités consistant à recueillir de l'information, ni de celle dont nous choisissons de les interpréter<sup>1</sup>.

Nous avons donc besoin d'une description des fonctions naturelles d'un système. Plus particulièrement, puisque nous nous intéressons à la fonction qu'un système de signes pourrait avoir, nous recherchons ce qu'un signe est *censé* signifier<sub>n</sub>, le

1. Je pense qu'une bonne partie de nos propos concernant les capacités représentationnelles des ordinateurs est de ce type assigné, et donc dérivé. Cela ne nous dit rien au sujet du pouvoir intrinsèque que possède une machine à représenter ou à représenter de manière erronée quoi que ce soit. En conséquence, cela ne nous dit rien du caractère cognitif de ses états internes. Cummins, je crois, saisit les choses parfaitement bien en distinguant la \*cognition (une version de la signification assignée) de la cognition authentique. Cf. Robert Cummins, *The Nature of Psychological Explanation*, Cambridge, MIT Press, 1983.

terme « censé » étant interprété en termes de la fonction de ce signe (ou de ce système de signes) dans l'économie cognitive propre à l'organisme. Nous voulons savoir comment le chien représente le contenu des bagages – ce que l'odeur qu'il flaire signifie<sub>f</sub> pour lui.

### 3. BESOINS

Afin de déceler des fonctions naturelles, on regardera de toute évidence du côté de systèmes biologiques possédant un ensemble d'organes, de mécanismes et de processus qui ont été développés (enrichis, préservés) *à cause* du rôle vital qu'ils ont joué dans l'adaptation d'une espèce à son environnement. Une fonction liée à la cueillette d'information, essentielle dans la plupart des cas à la satisfaction d'un besoin biologique, ne peut être réalisée adéquatement que dans un système capable d'occuper des états qui servent de signes naturels des conditions externes (et parfois aussi *d'autres* conditions internes). Pour que l'amoncellement de photorécepteurs que nous appelons la rétine remplisse sa fonction (peu importe comment nous l'appréhendons exactement), les divers états de ces récepteurs doivent signifier<sub>n</sub> quelque chose concernant le caractère et la distribution du champ optique d'un système. Ce que les divers états de ces récepteurs signifient<sub>f</sub> sera déterminé (conformément à  $S_f$ ) par deux choses : (1) ce que le système de récepteurs a pour fonction d'indiquer; (2) la signification<sub>n</sub> des divers états qui permettent au système de remplir cette fonction.

Afin d'illustrer la façon dont  $S_f$  est censée fonctionner, il sera utile de considérer des organismes simples ayant des besoins biologiques manifestes – soit une chose et un état sans quoi ils ne pourraient survivre. J'affirme que cette façon d'aborder le problème de la représentation erronée est commode parce qu'elle trouve son application la plus attrayante à propos de mécanismes cognitifs servant à remplir des besoins biologiques de base. La prise en considération de systèmes *primitifs* nous confère en outre l'avantage supplémentaire d'éviter le genre de circularité

qui risquerait d'affliger l'analyse si nous faisons appel à des types de « besoins » (le besoin d'un traitement de texte, par exemple) qui résultent de désirs (le désir de produire une copie au propre plus rapidement, disons). Nous ne pouvons pas faire appel aux désirs à ce stade-ci de l'analyse, puisqu'ils possèdent déjà le type de contenu représentationnel que nous cherchons à comprendre.

Certaines bactéries marines possèdent des aimants internes (appelés magnétosomes) qui fonctionnent comme les aiguilles d'une boussole, s'alignant (et alignant les bactéries par le fait même) en parallèle au champ magnétique de la Terre<sup>1</sup>. Puisque ces lignes magnétiques tendent vers le bas (en direction du Nord géomagnétique) dans l'hémisphère nord (et vers le haut dans l'hémisphère sud), les bactéries vivant dans l'hémisphère nord se déplacent, orientées par leurs magnétosomes, en direction du nord géomagnétique. Bien que la valeur de survie de la magnétotaxie (le nom de ce mécanisme sensoriel) ne soit pas manifeste, il est légitime de supposer que son rôle est de permettre aux bactéries d'éviter les eaux de surface. Puisque ces organismes ne peuvent vivre qu'en l'absence d'oxygène, un mouvement en direction du Nord diamagnétique les protège des eaux de surface riches en oxygène en les orientant vers les sédiments situés plus bas, lesquels en sont relativement dépourvus. Les bactéries vivant dans l'hémisphère Sud ont leurs magnétosomes inversés, ce qui leur permet de se déplacer en direction du Sud géomagnétique tout aussi avantageusement. Si l'on déplace une bactérie australe vers l'Atlantique Nord, elle va s'anéantir en se déplaçant vers le haut (en direction du Sud magnétique), et donc vers l'environnement toxique, riche en oxygène, des eaux de surface.

Si un aimant orienté en direction opposée à celle du champ magnétique de la Terre est placé à proximité de ces bactéries, elles peuvent être entraînées vers un environnement létal. Bien que je reviendrai sur ce point plus bas (afin de remettre en question ce raisonnement), il semble bien que nous ayons là un

1. La source de cet exemple se trouve dans R. P. Blakemore et R. B. Frankel (1981), « Magnetic Navigation in Bacteria », *Scientific American*, 245, p. 6.

cas vraisemblable de représentation erronée. Puisque dans l'habitat ordinaire des bactéries, l'orientation interne de leurs magnétosomes signifie<sub>n</sub> qu'il y a relativement peu d'oxygène dans *cette* direction, et puisque l'organisme a précisément besoin de cette information pour survivre, il semble raisonnable d'affirmer que ce mécanisme sensoriel a pour fonction de permettre la satisfaction de ce besoin, de transmettre cette information, d'indiquer que l'entourage dépourvu d'oxygène se trouve dans *cette* direction. Si c'est là ce qu'il est censé signifier<sub>n</sub>, c'est ce qu'il signifie<sub>f</sub>. En conséquence, en présence d'un aimant et conformément à S<sub>f</sub>, l'état sensoriel de l'organisme représente de manière erronée l'emplacement des eaux dépourvues d'oxygène.

Cela ne veut pas dire, bien entendu, que les bactéries ont la croyance qu'il y a peu ou pas d'oxygène dans *cette* direction. L'aptitude à représenter de manière erronée n'est qu'*un* aspect de l'intentionnalité, qu'*une* des propriétés qu'un système représentationnel doit posséder pour être un système de croyances. Pour être une croyance, un état représentationnel doit aussi manifester (entre autres choses) l'opacité particulière aux attitudes propositionnelles, et à moins qu'elle ne soit augmentée de quelque manière, la signification<sub>f</sub> ne comporte pas (encore) ce niveau d'intentionnalité. Notre projet, cependant, est plus modeste. Nous recherchons une forme naturalisée de représentation erronée, et bien que nous n'ayons pas encore une explication de la croyance *fausse*, nous avons, il semble bien, une explicitation naturaliste du *contenu* faux.

À part quelques fioritures terminologiques et une manière quelque peu différente d'aborder le problème près, rien de ce que j'ai dit jusqu'à présent n'est particulièrement original. Je n'ai fait que retracer les étapes, certaines d'entre elles importantes, que d'autres ont déjà parcourues. Je pense tout particulièrement à l'analyse de la représentation linguistique de Stampe, selon laquelle le contenu (éventuellement faux) d'une représentation est identifié à ce qui serait la cause de ce que cette représentation possède les propriétés qu'elle possède dans des conditions de

fonctionnement optimal<sup>1</sup>; au développement de certaines idées fonctionnelles, proposées par Enç afin d'expliquer l'intentionnalité des états cognitifs<sup>2</sup>; à l'application de notions téléologiques, avancée par Fodor afin de produire une sémantique pour son « langage de la pensée »<sup>3</sup>; et à la puissante analyse de la signification, que Millikan décrit en termes de cette variété de fonctions adéquates qu'un événement reproductible (tel un son ou un geste) peut avoir<sup>4</sup>. J'ai moi-même vaguement tenté d'exploiter des idées fonctionnelles dans mon analyse de la croyance en définissant le contenu sémantique d'une structure en termes de l'information qu'elle a été développée pour véhiculer (et qu'elle a donc pour fonction de véhiculer)<sup>5</sup>.

#### 4. L'INDÉTERMINATION DES FONCTIONS

Bien que cette approche du problème de la signification – et donc de la représentation erronée – ait été exploitée jusqu'à un certain point, certaines difficultés demeurent et font obstacle à ce qu'elle ne soit tenue ne serait-ce que pour une ébauche prometteuse, *a fortiori* d'une étude achevée, de la manière dont la nature commet une erreur.

Il y a tout d'abord le problème qui consiste à comprendre comment un système est à même de représenter de manière erronée quelque chose dont elle n'a pas biologiquement besoin. Si O n'a pas besoin de F (ou n'a pas besoin d'éviter F), aucun de ses systèmes cognitifs (suivant la présente approche) ne peut avoir pour fonction *naturelle* de lui signaler la présence

1. Dennis Stampe (1977), « Towards a Causal Theory of Linguistic Representation », *Midwest Studies in Philosophy*, 2, p. 42-63.

2. Berent Enç (1982), « Intentional States of Mechanical Devices », *Mind*, 91, p. 161-82. Enç identifie le contenu d'un état fonctionnel avec les (la construction des) propriétés de l'événement auquel le système a pour fonction de répondre.

3. Jerry Fodor (1986), « Why Paramecia don't have Mental Representations », *Midwest Studies in Philosophy*, 10, pp. 3-23.

4. Ruth Millikan (1984), *Language, Thought, and other Biological Categories*, Cambridge, MIT Press.

5. Fred Dretske (1981), *op. cit.*, troisième partie.

(l'absence, l'emplacement, l'approche, l'identité) de F. Et en l'absence de telles fonctions, il n'y a aucune possibilité de représenter de manière erronée quelque chose *en tant que* F. Quelque état interne pourrait toujours signifier<sub>n</sub> la présence d'un F (de la même façon que le système de détection de Fido signifie<sub>n</sub> que les bagages contiennent de la marijuana), mais cet état interne ne peut pas le signifier<sub>f</sub>. Ce que nous avons jusqu'à présent est une manière de comprendre comment un organisme peut représenter de manière erronée la présence de nourriture, d'un obstacle, d'un prédateur ou d'un partenaire (quelque chose qui est biologiquement nécessaire de se procurer ou d'éviter)<sup>1</sup>, mais nous ne comprenons toujours pas comment nous pouvons représenter de manière erronée quelque chose en tant que, disons, ouvre-boîte, raquette de tennis, tulipe ou valet de carreau. Même si nous supposons que notre système nerveux est suffisamment sophistiqué pour indiquer (dans des conditions normales) la présence de tels objets, ces états neuronaux n'ont certainement pas pour fonction *naturelle* de signaler la présence – et encore moins la présence de sortes spécifiques – d'ustensiles de cuisine, d'équipement sportif, de fleurs et de cartes à jouer.

Bien qu'il s'agisse là d'une difficulté redoutable, elle n'est *pas*, je pense, insurmontable. Car manifestement, un système cognitif peut être formé afin de servir, et ainsi avoir pour fonction naturelle de servir, un certain besoin biologique sans pour autant que ses activités représentationnelles (même erronées) ne soient confinées à ces besoins. Dans le but d'identifier son prédateur naturel, un organisme pourrait former des détecteurs de couleur, de forme, et de mouvement d'un pouvoir discriminatoire considérable. Muni alors de cette capacité de différencier les couleurs, les formes et les mouvements, l'organisme acquiert, à titre d'avantage subsidiaire, pourrait-on dire, la capacité d'identifier (et donc, de mal identifier) des choses dont il n'a pas biologiquement besoin. L'organisme peut n'avoir aucun besoin

1. Quelque chose pour lequel il y a, pour parler comme Dennett naguère, une « continuation efférente appropriée ». Voir Daniel Dennett (1969), *Content and Consciousness*, London, Routledge & Kegan Paul.

de feuilles vertes, mais son besoin de fleurs roses a conduit à la formation d'un système cognitif dont les états variés peuvent, à cause de leur signification<sub>f</sub> reliée au besoin en question, signifier<sub>f</sub> la présence de feuilles vertes. Bien que n'ayant aucun besoin de feuilles vertes, l'organisme a peut-être même développé un goût pour ces feuilles, et donc un moyen de les représenter à l'aide d'éléments possédant déjà une signification<sub>f</sub>.

Il y a cependant une objection plus sérieuse à cette approche du problème de la représentation erronée. Considérons à nouveau les bactéries. Nous avons dit que leur système magnétotactique avait pour fonction d'indiquer l'emplacement d'environnements dépourvus d'oxygène. Mais pourquoi décrire la fonction de ce système de cette manière? Pourquoi ne pas dire que ce système a pour fonction d'indiquer la direction du Nord géomagnétique? Plus modestement, nous devrions peut-être même assigner à ce détecteur la fonction d'indiquer l'emplacement (la direction) du Nord magnétique (et pas nécessairement du Nord géomagnétique). Ce mécanisme sensoriel élémentaire fonctionne, après tout, parfaitement bien lorsque, influencé par un aimant, il entraîne son détenteur vers un environnement toxique. Lorsque cela se présente, il y a bien entendu *quelque chose* qui ne va pas, mais je ne vois aucune raison de faire porter le blâme sur le mécanisme sensoriel, aucune raison de dire qu'il ne remplit pas sa fonction. On pourrait tout aussi bien reprocher à une jauge d'essence de ne pas remplir sa fonction lorsque le réservoir est rempli d'eau (et que la conductrice est conséquemment induite en erreur concernant la quantité d'*essence* qui lui reste). En de pareilles circonstances anormales, l'instrument s'acquitte de ses tâches de façon parfaitement satisfaisante — c'est-à-dire en indiquant la quantité de liquide dans le réservoir. Ce qui a fait défaut est quelque chose dont l'instrument lui-même ne saurait être tenu responsable, à savoir une brèche au sein des corrélations normales (entre la quantité de liquide et la quantité d'essence dans le réservoir) qui permettent à la jauge de servir de jauge d'*essence*, qui lui permettent (lorsque les conditions sont normales) de signifier<sub>n</sub> qu'il y a de l'essence dans le

réservoir. De la même manière, rien ne fait défaut dans votre système perceptif lorsque vous consultez une horloge qui prend du retard et vous trompez donc sur l'heure qu'il est. Vos yeux ont pour fonction de vous dire ce que *l'horloge indique*; *l'horloge a pour fonction d'indiquer l'heure qu'il est*. Saisir les choses correctement concernant ce que vous avez besoin de connaître est souvent une responsabilité *partagée*. Vous devez saisir G correctement, et G doit saisir F correctement. En conséquence, même si c'est de F que vous avez besoin, ou que vous avez besoin de connaître, votre système perceptif peut n'avoir pour fonction que de vous informer à propos de G.

Si nous appliquons ce raisonnement au système sensoriel de la bactérie, sa fonction est alors d'aligner l'organisme au champ magnétique prévalent. C'est la tâche du Nord magnétique, en quelque sorte, que d'être la direction conduisant à des eaux dépourvues d'oxygène. En déplaçant une bactérie boréale dans l'hémisphère Sud, nous pouvons certes tout bouleverser, mais *pas* parce qu'un déplacement hémisphérique entraîne une désorientation *sensorielle*. Non, le système magnétotactique fonctionne tel qu'il est censé fonctionner, tel qu'il a été (vraisemblablement) développé au fil de l'évolution afin de fonctionner. Tout au plus, on pourrait affirmer qu'il y a une certaine erreur *cognitive* (de sa condition sensorielle, la bactérie « infère » à tort que les eaux dépourvues d'oxygène se trouvent dans cette direction). Mais ce genre de réponse est circulaire en ce qu'elle présuppose que l'organisme possède déjà la capacité conceptuelle ou représentationnelle de représenter quelque chose *en tant que* direction des eaux dépourvues d'oxygène. Notre question est de savoir *si* oui ou non l'organisme possède cette capacité et, le cas échéant, d'en connaître la provenance<sup>1</sup>.

1. Fodor (dans une version manuscrite de « Why Paramecia don't have Mental Representations »; voir Jerry Fodor *Psychosemantics*, Cambridge, MIT Press, 1987) distingue des organismes pour lesquels une théorie représentationnelle de l'esprit n'est pas appropriée (les paramécies, par exemple) de ceux pour lesquels elle l'est (nous, par exemple), en termes de l'habileté de ces derniers à répondre à des propriétés non nomiques d'un stimulus (des propriétés qui ne sont pas détectables par un transducteur). Nous sommes capables, contrairement aux

Les bactéries boréales, il est vrai, n'éprouvent pas le besoin de vivre dans un climat nordique en tant que tel. Et donc, décrire la fonction des détecteurs d'une bactérie en termes du rôle qu'ils jouent dans l'identification du Nord géomagnétique n'est pas la manière de décrire de manière à révéler *comment* cette fonction est reliée à la satisfaction de ses besoins. Mais nous n'avons pas à décrire la fonction d'un mécanisme en termes des besoins biologiques ultimes de son détenteur<sup>1</sup>. Le cœur a pour fonction d'assurer la circulation sanguine. Précisément *pourquoi* le sang a besoin de circuler peut être un mystère.

Alors la question épineuse est la suivante : *étant donné* qu'un système a besoin de F, et *étant donné* qu'un mécanisme M rend l'organisme capable de détecter, identifier ou reconnaître F, *comment* le mécanisme s'acquitte-t-il de cette fonction ? Est-ce en représentant les F avoisinants en tant que F avoisinants, ou est-ce, peut-être, en les représentant simplement comme des G avoisinants, laissant à la nature (la corrélation entre F et G) le

paramécies, de représenter quelque chose comme chemise chiffonnée, disons, et être une chemise chiffonnée n'est pas une propriété projectible. Dans cet article, Fodor ne s'intéresse pas à la question de savoir d'où provient cet extraordinaire pouvoir représentationnel (il suggère que ce pouvoir requiert des capacités inférentielles). Il ne s'y intéresse que dans la mesure où ce pouvoir lui sert à nous distinguer d'une variété d'autres systèmes perceptifs et quasi cognitifs.

Je suis d'accord avec Fodor concernant l'importance et la pertinence de cette distinction, mais mon intérêt ici est de comprendre *comment* un système peut acquérir le pouvoir de représenter quelque chose de cette façon. Le pouvoir de représenter quelque chose en tant que chemise chiffonnée (où cela implique la capacité corrélatrice de sa représentation erronée) n'est certainement pas inné.

1. Enç (1982), p. 168, dit qu'un photorécepteur dans la drosophile a pour fonction de rendre la mouche capable d'accéder à des lieux humides (en vertu de la corrélation entre les lieux sombres et les lieux humides). Je n'ai aucune objection à décrire les choses de cette manière. Toutefois, la question demeure : *comment* le photorécepteur remplit-il cette fonction ? Nous pouvons répondre à cette question sans supposer qu'il y ait de mécanisme dans la mouche ayant pour fonction d'indiquer le degré d'humidité. Le mécanisme sensoriel peut s'acquitter de cette fonction pour peu qu'il y ait quelque chose pour indiquer la luminosité – c'est-à-dire un photorécepteur. Cela rendra la mouche capable d'accéder à des lieux humides. De la même manière, le sens magnétotactique des bactéries les rend capables (et, disons, a pour fonction de les rendre capable) d'éviter les eaux riches en oxygène. Mais il le fait (pourrait-on argumenter) à l'aide d'un détecteur qui indique, et a pour fonction d'indiquer, la direction du champ magnétique.

pour satisfaire ses besoins ? Décrire un mécanisme cognitif en tant que détecteur de F (et, donc, comme un mécanisme jouant un rôle vital pour la satisfaction des besoins d'un organisme) ne nous fournit pas encore la trame fonctionnelle à l'aide de laquelle le mécanisme s'acquitte de sa tâche. Tout ce que nous savons, lorsque nous savons que O a besoin de F et que M rend O capable de détecter F, est que M *ou bien* signifie<sub>f</sub> la présence de F, *ou bien* signifie<sub>f</sub> la présence de G, où G est, dans l'environnement naturel de O, un signe naturel de la présence de F (où G signifie<sub>n</sub> F)<sup>1</sup>. Si j'ai besoin de vitamine C, on ne devrait pas automatiquement attribuer à mon système perceptif-cognitif la capacité de reconnaître les objets *en tant qu'ils* contiennent de la vitamine C (en tant qu'ils signifient<sub>f</sub> qu'ils contiennent de la vitamine C) simplement parce qu'il ne fournit pas l'information requise pour satisfaire mon besoin. Représenter les choses en tant qu'oranges et citrons fera très bien l'affaire.

Le problème auquel nous sommes confrontés n'est autre que celui de rendre compte des pouvoirs de représentation erronée que possède un système *sans*, pour ce faire, *exagérer* de manière fictive les fonctions naturelles de ce système. Nous avons besoin d'un principe [*principled way*] nous permettant d'établir ce qu'est la fonction naturelle d'un mécanisme, ce que ses divers états non seulement *signifient<sub>n</sub>*, mais ce qu'ils signifient<sub>f</sub>. Il semble un peu forcé (de mon point de vue, du moins) de décrire le mécanisme sensoriel des bactéries comme indiquant, et ayant

1. Pour parler comme Fodor (voir Jerry Fodor, (1990), « Psychosemantics : Or where do Truth Conditions come from », *Mind and Cognition*, sous la direction de W. Lycan, Oxford, Basil Blackwell), ceci n'est qu'une façon de dire qu'identifier la sémantique de M (quelque représentation mentale) avec les conditions entrantes (relatives à un ensemble de conditions de normalité) laisse tout de même un jeu. Nous pouvons dire que la condition entrante est l'absence (la présence) d'oxygène *ou* une orientation spécifique du champ magnétique. Faire appel à l'histoire sectionnelle de ce mécanisme ne pourra décider *quelle* est la spécification appropriée des conditions intrantes – et donc, ne nous dira pas si les bactéries sont capables ou non de représenter de manière erronée quoi que ce soit. Fodor, je pense, constate ce résidu d'indétermination et fait la remarque allusive (n. 9) que ce problème est analogue à celui qui consiste, pour les théories de la perception, à spécifier l'objet perceptuel.

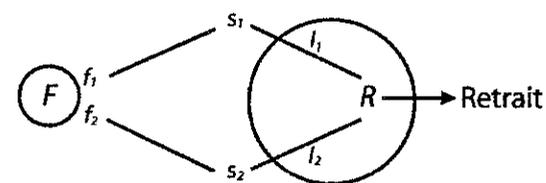
la fonction d'indiquer, l'emplacement d'oxygène. Car cela revient à dire que le mécanisme ne remplit pas sa fonction dans des conditions trompeuses (en présence d'un aimant, par exemple). Il s'agit, après tout, d'un détecteur *magnétotactique*, et non pas *chimiotactique*. Mais si nous choisissons de décrire la fonction de ce senseur de cette manière plus modeste, nous n'avons plus dès lors l'exemple d'un système susceptible d'avoir une représentation erronée. Une bactérie boréale (déplacée dans l'hémisphère Sud) ne pourra faussement représenter quoi que ce soit lorsque, guidée par son détecteur magnétotactique, elle se déplacera vers le haut (vers le nord géomagnétique), en direction de l'environnement léthal des eaux de surface. L'alignement de ses magnétosomes signifiera<sub>n</sub> ce qu'il a toujours signifier<sub>n</sub>, ce qu'il a pour fonction de signifier<sub>n</sub>, ce qu'il est censé signifier<sub>n</sub> : à savoir que le Nord magnétique se trouve dans cette direction. La catastrophe peut être attribuée à l'environnement anormal. Nous ne pouvons pas non plus récupérer un quelconque reste de capacité de représenter de manière erronée en supposant que la bactérie, influencée par l'aimant, représente faussement la direction du Nord géomagnétique. Car, encore une fois, le même problème se présente : pourquoi supposer que le mécanisme a pour fonction d'indiquer la direction du nord géomagnétique, et non pas, tout simplement, la direction du champ magnétique prévalent ? Si nous décrivons la fonction seulement de cette dernière façon, il devient impossible de duper l'organisme, impossible de faire en sorte qu'il représente de manière erronée quoi que ce soit. Car ses états internes signifient<sub>f</sub> seulement que le champ magnétique se trouve dans *cette* direction, et (comme une boussole) cela est toujours juste.

### 5. DÉTERMINATION FONCTIONNELLE

Dans le but d'éclaircir la question, j'ai jusqu'à présent limité la discussion à des organismes simples possédant des capacités représentationnelles élémentaires. Il n'est donc pas surprenant de ne trouver à ce niveau aucune capacité de représentation

erronée qui soit claire et non équivoque. Car ce pouvoir – de même, vraisemblablement, que la capacité qui en dépend de former des croyances – requiert un certain niveau de complexité des moyens de traitement de l'information dont un système dispose. Quelque part entre l'organisme unicellulaire et l'être humain, nous franchissons ce seuil. Le but de cette dernière partie est de décrire la nature de ce seuil, de décrire le *genre* de complexité responsable de la capacité de représentation erronée manifestée par les organismes supérieurs.

Supposons qu'un organisme (à la différence de notre bactérie) ait deux façons de détecter la présence d'une certaine substance toxique F. Ceci peut être parce que l'organisme est pourvu de deux modes sensoriels, chacun étant (à sa façon) sensible à F (ou à un quelconque signe naturel de F spécifique à l'un ou l'autre des modes sensoriels), ou parce qu'un mode sensoriel unique exploite divers signes (ou indices) extérieurs de F. On peut illustrer ce dernier point en considérant comment, par exemple, nous pourrions visuellement identifier un chêne de l'une de deux manières : par le motif caractéristique de ses feuilles (en été), ou par la texture et le motif caractéristiques de son écorce (en hiver). Nous avons donc deux états ou conditions internes,  $I_1$  et  $I_2$ , chacun étant le produit d'une chaîne différente d'événements antérieurs qui sont des signes naturels de la présence de F. Chacun des deux états signifie<sub>n</sub> la présence de F. Supposons, en outre, que puisque l'organisme a besoin d'échapper à la substance toxique F, ces états internes soient attelés à un troisième état, appelons-le R, lequel déclenche ou occasionne un comportement de fuite. Le schéma 1 réunit tous les détails pertinents. R, bien entendu, est aussi un signe naturel de F. Dans des conditions normales, R ne survient qu'en présence de F.  $f_1$  et  $f_2$  sont des propriétés typiques de F ordinaires.  $s_1$  et  $s_2$  sont des stimuli proximaux.



Si maintenant nous exposons le système à un quelconque ersatz de  $F$  (analogue à l'aimant dans le cas de la bactérie), quelque chose qui affiche certaines des propriétés de  $F$  (faisons), nous déclenchons une chaîne d'événements ( $s_1, I_1, R$ , et la fuite) qui normalement ne survient, et n'est vraiment appropriée, qu'en présence de  $F$ . Si nous considérons l'état interne  $R$  et demandons ce qu'il signifie<sub>f</sub> dans ces conditions trompeuses, nous nous voyons incapables de dire (comme nous le pouvions dans le cas des bactéries) qu'il signifie<sub>f</sub> quoi que ce soit (c'est-à-dire une des conditions proximales) si ce n'est  $F$  lui-même. Bien que  $s_1$  déclenche (par l'entremise de  $I_1$ ) l'occurrence de  $R$ ,  $R$  ne signifie<sub>n</sub> pas (et donc, ne peut signifier<sub>f</sub>) que  $s_1$  (ou  $f_1$ ) survient.  $R$  est comparable à une ampoule qui est reliée à des interrupteurs branchés en parallèle et qui s'allume pour peu que l'un ou l'autre soit activé. Lorsque l'ampoule s'allume, cela ne signifie<sub>n</sub> pas que l'interrupteur n 1 est actionné, même si c'est l'activation de cet interrupteur qui en est la cause. Cet événement ne peut le signifier<sub>n</sub>, parce qu'il n'y a pas de corrélation régulière entre l'ampoule qui s'allume et l'activité de l'interrupteur n 1 (dans 50 pour cent des cas, c'est l'activation de l'interrupteur n 2 qui en est la cause).

Si nous attribuons au système détecteur décrit ci-dessus la fonction de rendre l'organisme capable de détecter  $F$ , les multiples façons dont il dispose pour ce faire ont alors pour conséquence que certains états internes ( $R$ , par exemple) peuvent indiquer (et donc signifier<sub>f</sub>) la présence de  $F$  sans indiquer quoi que ce soit au sujet des conditions intermédiaires (c'est-à-dire  $f_1$  ou  $s_1$ ) qui lui « disent » qu'il est en présence de  $F$ . Notre problème avec les bactéries était de trouver un moyen de faire en sorte que l'orientation de leurs magnétosomes signifie<sub>f</sub> que les eaux dépourvues d'oxygène se trouvaient dans une certaine direction sans pour autant exclure *arbitrairement* la possibilité qu'elle puisse signifier<sub>f</sub> que le champ magnétique était aligné dans cette direction. Nous sommes maintenant en mesure de voir qu'avec les ressources multiples illustrées par le schéma 1, cette possibilité peut être non arbitrairement exclue.  $R$

ne peut pas signifier<sub>f</sub> que  $f_1$  ou  $s_1$  survient, parce que même dans les conditions optimales, il ne peut pas le signifier<sub>n</sub>. Nous pouvons donc prétendre avoir trouvé un cas de représentation erronée non dérivé (c'est-à-dire un cas où  $R$  signifie<sub>f</sub> la présence de  $F$  même en son absence) que l'on ne saurait écarter en décrivant ce que  $R$  signifie<sub>f</sub> de manière à éliminer toute apparence de représentation erronée. Le danger d'une inflation des significations<sub>f</sub> possibles, résultant de la variété de façons dont la fonction naturelle d'un système pourrait être décrite, a été neutralisé.

Nous ne sommes tout de même pas *forcés* de tenir ce cas de représentation erronée pour authentique, dira-t-on, si nous sommes disposés à reconnaître que  $R$  possède une signification<sub>n</sub> disjonctive. Lorsque l'ampoule (reliée à des interrupteurs branchés en parallèle) s'allume, l'événement ne signifie<sub>n</sub> certes pas qu'un interrupteur en particulier est activé, mais il indique toutefois qu'un des interrupteurs l'est. De la même manière, on pourrait dire que bien que le mécanisme comportant  $R$  comme état final ait pour fonction d'alerter l'organisme de la présence de  $F$ , il ne remplit cette fonction que parce que  $R$  indique, et a pour fonction d'indiquer, l'occurrence d'une certaine condition disjonctive – nommément,  $f_1$  ou  $f_2$  (ou  $s_1$  ou  $s_2$ ). Notre organisme hypothétique fuit  $F$  à tort *non pas* parce qu'il représente de manière erronée l'ersatz de  $F$  en tant que  $F$ , mais parce que ce qu'il indique correctement (c'est-à-dire que l'ersatz de  $F$  est  $f_1$  ou  $f_2$ ) n'est plus corrélé de manière normale avec le fait que quelque chose soit  $F$ .

Peu importe la polyvalence que nous pourrions concevoir pour un système détecteur, peu importe le nombre de voies d'accès de l'information dont nous pourrions doter un organisme, il sera toujours possible de décrire sa fonction (et donc la signification<sub>f</sub> de ses divers états) comme étant la détection de quelque propriété hautement disjonctive de l'intrant proximal. À tout le moins, cela sera toujours possible *si* nous avons un ensemble déterminé d'éléments disjoints sur lequel nous rabattre.

Supposons cependant que nous ayons un système capable d'une forme d'apprentissage associatif. Supposons, pour le dire autrement, que suite à une exposition répétée à SC (un stimulus conditionné) en présence de F, un changement survienne. R (et donc, le comportement de fuite) peut maintenant être déclenché par l'occurrence de SC seul. Qui plus est, il est clair qu'il n'y a pratiquement aucune limite quant au type de stimulus pouvant acquérir cette efficacité « déplacée » à déclencher R et le comportement de fuite qui s'ensuit. Presque n'importe quel stimulus peut devenir un SC, prenant ainsi le « contrôle » de R, en fonctionnant (au niveau de l'« expérience » de l'organisme) en tant que signe de F.

Nous avons dès lors un mécanisme cognitif qui non seulement transforme une variété d'intrants sensoriels ( $s_i$ ) en un seul état qui détermine l'extrait (R), mais est aussi capable de modifier le caractère des divers intrants responsables de l'extrait au fil du temps. Si nous nous en tenons aux intrants sensoriels (le  $s_1$  du schéma 1), R signifie<sub>n</sub> une chose au temps  $t_1$  ( $s_1$  ou  $s_2$ , par exemple), autre chose à  $t_2$  ( $s_1$  ou  $s_2$ , ou, suite à un apprentissage,  $sc_3$ , par exemple), et autre chose encore à un moment ultérieur. Ce que R signifie<sub>n</sub> précisément dépendra de l'histoire de l'apprentissage de l'individu – de quel  $s_1$  est devenu  $sc_1$  pour lui. Il n'y a pas de signification<sub>n</sub> qui soit temporellement invariante pour R; donc, rien qu'il ne pourrait avoir, au fil du temps, la fonction d'indiquer. En termes du  $s_1$  qui produit R, R ne peut avoir de signification<sub>f</sub> temporellement invariante.

Tout au long de ce processus, R continue d'indiquer la présence de F, bien entendu. Il le fait parce que, par hypothèse, tout nouvel  $s_i$  auquel R est conditionné est un signe naturel de F. L'apprentissage est un processus au cours duquel les stimuli indiquant la présence de F sont à leur tour indiqués par un certain état interne de l'organisme (R dans le cas présent). En conséquence, si nous nous représentons ces mécanismes cognitifs comme ayant une fonction temporellement invariante (ce qui découle de leur entretien continu – et, suite à l'apprentissage, certes plus efficient – du besoin correspondant), nous devons

alors nous représenter cette fonction non pas comme indiquant la nature des conditions proximales (ou même distales) qui déclenchent des réponses positives ( $s_1$  et  $f_1$ ), mais comme indiquant la condition (F) dont ces divers stimuli sont des signes. Le mécanisme qu'on vient de décrire a pour fonction naturelle d'indiquer la présence de F. En conséquence, l'occurrence de R signifie<sub>f</sub> la présence de F. Elle ne signifie<sub>f</sub> pas que  $s_1$  ou  $s_2$  ou...  $s_x$  est réalisé, même si, à toute phase donnée du développement, elle le signifiera<sub>n</sub> pour une valeur définie de  $x$ .

Un système comportant ce niveau de complexité, ayant non seulement de multiples voies d'accès à ce qu'il a besoin de connaître, mais aussi les ressources nécessaires à l'accroissement de ses moyens de recueillir de l'information, possède, et c'est là ma thèse, un pouvoir authentique de représentation erronée. Lorsqu'une brèche au sein de la chaîne normale des signes naturels intervient, lorsque, disons,  $sc_7$  (un signe de F appris) survient en des circonstances où il ne signifie<sub>n</sub> pas la présence de F (de la même manière que l'horloge défectueuse ne signifie<sub>n</sub> pas qu'il est 3h 30), R signifie<sub>f</sub> toujours (quoi qu'il ne signifie<sub>n</sub> pas, bien entendu) la présence de F. Il le signifie<sub>f</sub> parce que c'est là ce qu'il est censé signifier<sub>n</sub>, ce qu'il a pour fonction naturelle de signifier<sub>n</sub>, et qu'aucune autre condition qu'il serait susceptible de signifier<sub>f</sub> n'est disponible<sup>1</sup>.

FRED DRETSKE

*Traduit de l'anglais par Robert Lantin*

1. Je suis reconnaissant à Berent Eng, Dennis Stampe et Jerry Fodor pour leurs critiques à la fois constructives et destructives des versions antérieures de ce texte.