UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LE PASSAGE À L'EXPERTISE MOTRICE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN PHILOSOPHIE

PAR

MARTIN RATTE

MAI 2013

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens d'abord à remercier mon directeur de thèse, Pierre Poirier. Le chemin que j'ai dû parcourir avant de terminer ma thèse a été long et ardu et, sans son aide, je ne serais probablement jamais arrivé à bon port. Serge Robert et Luc Faucher, m'ont également aidé de manière significative. Plus précisément, je tiens à les remercier pour les conseils qu'ils m'ont donnés alors que ma thèse n'était encore qu'un projet. Enfin, je remercie le professeur Pierre Livet pour m'avoir fait de précieux commentaires lors de l'évaluation.

TABLE DES MATIÈRES

LISTE DES FIGURES viii

LISTE DES TABLEAUX ix

RÉSUMÉ x

CHAPITRE I
INTRODUCTION 1

1.1 Un modèle causal de l'action 4

1.2 La thèse que je soutiens 12

1.3 Division et subdivision de mon texte 17

CHAPITRE II
L'ARCHITECTURE COGNITIVE À LA BASE DU PLAN D'ACTION 21

2.1 Les intentions comme états mentaux à part entière 22

2.1.1 Les représentations et les attitudes 22

2.1.2 La non réductibilité des intentions 24

2.2 L'architecture cognitive à la base du plan d'action 30

2.2.1 Le raisonneur procédant des fins aux moyens 32

2.2.2 Les autres composantes de l'architecture cognitive 38

2.3 Comment expliquer le passage à l'expertise? 45

2.3.1 Pourquoi l'agent planificateur n'est pas un agent expert? 45

2.3.2 La « réactivité » de notre architecture cognitive 47

2.4 Résumé 48

CHAPITRE III
LA PERCEPTION DES AFFORDANCES 53

3.1 Les affordances 53

3.1.1 La position de Turvey (1992) 55

3.1.2 La position de Stoffregen (2003) 59

3.1.3 Les objections de Stoffregen (2003) 60

3.2 La perception des affordances 63

3.2.1 La structure optique peut-elle transmettre de l'information sur l'affordance? 64

3.2.2 En quoi consiste la structure optique qui spécifie une affordance? 72

3.2.3 La relation à l'effectivité dans la perception d'une affordance 74

3.3 Résumé 77

CHAPITRE IV
UNE DÉFENSE DE LA PERCEPTION DIRECTE 81

4.1 Le domaine des propriétés directement perceptibles 81

4.2 La 1ère objection de Fodor et Pylyshyn 87

4.3 La 2ème objection de Fodor et Pylyshyn : l'objection du « stimulus effectif » 92

4.4 Résumé 96

CHAPITRE V
LE RÉSEAU DE NEURONES À LA BASE DE LA PERCEPTION DES AFFORDANCES .. 100

5.1 Une introduction partielle aux réseaux de neurones 100

5.2 Le réseau de neurones à la base de la perception des affordances 105

5.2.1 Un réseau de neurones doté d'une mémoire dynamique 105

5.2.2 Jordan ou Elman? 110

5.2.3 La détection d'un invariant d'invariants 115

5.3 Résumé 116

CHAPITRE VI
LES VARIABLES OPTIQUES ET LE CONTRÔLEUR MOTEUR 119

6.1 Les paramètres de contrôle des muscles 119

6.2 Les variables optiques de contrôle 123

6.3 Le contrôleur moteur 126

6.4 Résumé 132

2ÈME PARTIE : LES MÉCANISMES D'APPRENTISSAGE 133

CHAPITRE VII
L'APPRENTISSAGE MOTEUR 138

7.1 Aperçu sur la théorie des systèmes dynamiques 139

7.1.1 Le comportement des systèmes dynamiques 140

7.1.2 Le mécanisme qui est à la base du comportement d'un système dynamique 145

7.2 Le mécanisme d'apprentissage inspiré de Ashby 152

7.2.1 Le système R interagit avec l'environnement (les variables optiques) 152

7.2.2 La paramétrisation du système R 153

7.2.3 La variable d'appréciation 154

7.2.4 L'action de la variable d'appréciation sur les paramètres S 156

7.3 Résumé 164

CHAPITRE VIII
L'APPRENTISSAGE D'UNE BONNE POLITIQUE PERCEPTUELLE 165

8.1 Le problème de l'apprentissage par renforcement 166

8.2 La méthode de programmation dynamique 172

8.2.1 L'évaluation de la politique actuelle 173

8.2.2 Vers une politique optimale 179

8.3 La méthode Monte-Carlo 180

8.3.1 L'évaluation Monte-Carlo de la politique 181

8.3.2 Le problème lié à l'amélioration de la politique et la solution à ce problème 184

8.4 La méthode des différences temporelles 186

8.5 Retour sur notre propre problème 189

8.6 Résumé 193

CHAPITRE IX
APPRENDRE À RÉAGIR 196

9.1 Le mécanisme d'apprentissage 197

9.2 Quelques objections 205

CONCLUSION 210

RÉFÉRENCES 223

LISTE DES FIGURES

Figure Page

1.1 Les relations causales entre les différents éléments qui déterminent l'action 6

1.2 Le contrôle des perceptions par les intentions chez le novice 15

1.3 Le contrôle des perceptions par les intentions chez l'expert 16

2.1 L'architecture cognitive 30

2.2 Un arbre de recherche 37

5.1 Les réseaux de Elman et Jordan 104

5.2 L'erreur associée à l'output 107

5.3 L'erreur associée à l'output 109

6.1 Le flux optique lors de la locomotion 123

6.2 Les boucles de feedback interne et externe 127

6.3 Le prédicteur de Smith 129

7.1 L'interaction entre le système R et l'environnement 153

7.2 L'action des paramètres S sur le système R 154

7.3 L'ajout de la variable d'appréciation 155

7.4 L'action de la variable d'appréciation sur les paramètres S 157

7.5 Portrait de phase d'un système ayant une variable « step-function » 160

7.6 Les portraits de phase correspondant aux différentes valeurs du système S 161

8.1 Exemple de la méthode itérative d'évaluation 178

LISTE DES TABLEAUX

Tableau Page

5.1 Les vecteurs représentant les lettres 110

RÉSUMÉ

L'objectif de la présente thèse est d'expliquer le passage de novice à expert en matière d'action motrice. L'expert et le novice basent tous deux leurs actions sur une structure « cognitive ». Dans mon modèle de l'action motrice, la structure « cognitive » de l'expert et la structure « cognitive » du novice partagent les mêmes éléments. Ces éléments sont 1) des perceptions de variables optiques de contrôle, 2) des perceptions d'affordances, 3) des intentions (un plan) et 4) un contrôleur moteur. Dans la première partie de ma thèse, je m'applique à définir ces quatre éléments. Lorsque l'agent passe de novice à expert, ces éléments subissent des changements, et la présente thèse a tenté de défendre ces changements. Ces changements ont pour conditions de possibilité des processus d'apprentissage. Donc, pour défendre ces changements, j'ai tenté, dans la seconde partie de ma thèse, de trouver les processus d'apprentissage qui sont à leur origine. Le premier changement dont il a été question est un changement moteur. Lorsque l'agent est novice, les commandes motrices du contrôleur moteur sont inappropriées, tandis que lorsqu'il est expert, ces commandes motrices sont appropriées. Le mécanisme d'apprentissage dont je me suis servi pour expliquer ce changement a été inspiré par le texte de Ashby, *Design for a Brain*. Le deuxième changement dont il a été question est un changement perceptuel. Lorsqu'il est novice, l'agent ne perçoit pas les bonnes affordances et les bonnes variables optiques de contrôle, tandis que, une fois expert, il les perçoit. Pour rendre compte de ce changement, je me suis inspiré de l'apprentissage par renforcement tel qu'il a été défini par Sutton et Barto (1998). Le troisième et dernier changement dont il a été question consiste en une diminution du plan. Lors de la diminution du plan, l'agent devient de plus en plus réactif. Ce changement est rendu possible par une « fusion » des politiques perceptuelles. Et, pour être optimale, la « fusion » des politiques perceptuelles doit reposer sur un apprentissage par renforcement, le même que celui qui est à l'origine de notre capacité à percevoir les bonnes affordances et les bonnes variables optiques de contrôle.

Mots-clés : Affordance; plan; réactivité; variable optique de contrôle; contrôleur moteur.

CHAPITRE I

INTRODUCTION

Le problème général que le présent travail se pose est le suivant : Qu'est-ce qui caractérise le passage de novice à expert en matière d'action motrice? Pour répondre à cette question, je dois d'abord prendre le temps de développer un modèle *causal* de l'action. En effet, on peut supposer que le passage d'une action de novice à une action d'expert résulte d'un changement de la *cause* interne de l'action. Lorsque l'agent est novice, la cause interne est dans un état donné et cet état cause une action de novice, tandis que, une fois l'agent devenu expert, la cause interne change d'état et ce nouvel état cause une action experte. Il est donc effectivement primordial, avant de répondre à la question qui vient d'être énoncée, de se doter d'un modèle *causal* de l'action. Une fois que nous aurons en notre possession ce modèle causal, nous connaîtrons les éléments qui causent l'action et, pour répondre à notre question, il nous suffira de déterminer les changements qui touchent ces éléments lorsque l'agent passe de novice à expert. Mais avant de tracer les grandes lignes de mon modèle causal de l'action, définissons d'abord le type d'action motrice sur lequel porte ma thèse.

L'action motrice, contrairement à l'action de raisonner ou de planifier, est une action qui implique un mouvement. Mais l'action motrice n'est pas un simple mouvement. Quelqu'un peut déplacer mon bras, mais je ne peux pas dire pour autant que j'ai exécuté l'action de bouger le bras. Il semble que, pour qu'il y ait une action, il faille que le mouvement soit causé, non pas par une cause extérieure à l'agent, mais par une cause interne à l'agent. Traditionnellement, on a considéré que cette cause interne était constituée de la croyance et du désir de l'agent. Donc, selon cette perspective, qui est défendue par plusieurs, dont Davidson (1980), l'action est simplement un mouvement qui est causé par une cause interne appropriée (une croyance et un désir appropriés). Comme la précédente, cette perspective sur

l'action réduit l'action à un simple mouvement, sauf qu'elle ajoute que, pour qu'il soit une action, le mouvement doit être impliqué dans une relation causale avec la croyance et le désir de l'agent. Cette analyse de l'action ne fait pas l'unanimité. Par exemple, bien que Dretske (1988) soutienne que l'action motrice comprenne comme partie le mouvement qui est causé par une cause interne appropriée, il rajoute que l'action motrice comprend aussi cette cause interne comme partie. Ainsi, chez Dretske, l'action motrice est comprise comme un processus, comme la causation ou la production d'un mouvement par une cause interne. L'adoption de l'une ou l'autre position a des conséquences importantes sur mon objectif immédiat, qui est de développer un modèle causal de l'action. En effet, selon que l'on adopte la position de Dretske ou de Davidson, ce projet prend deux sens complètement différents. Si je définis l'action comme le fait Davidson, alors ce projet consiste à déterminer les causes internes du mouvement, tandis que, si je définis l'action comme le fait Dretske, mon projet consiste à déterminer les causes d'un processus, c'est-à-dire, plus précisément, à déterminer les causes de la production ou causation du mouvement par une cause interne. Or, ci-dessus, lorsque je faisais allusion au projet de trouver un modèle causal de l'action, je n'avais pas en tête de trouver les causes d'un processus qui comprend un mouvement comme partie finale, mais bien de trouver les causes internes d'un mouvement. J'avais donc en tête la notion davidsonienne d'action. Donc, pour la suite des choses, j'utiliserai le terme « action » avec le sens que Davidson lui donne. Autrement dit, l'action sera pour moi un mouvement entrant en relation avec une cause interne appropriée[1].

Avant de définir mon modèle causal du mouvement et, donc, avant de déterminer les causes internes du mouvement, une dernière chose devrait être dite à propos de l'action motrice. L'action motrice qui m'intéresse est une action sensorimotrice. Par le terme « action sensorimotrice », on peut penser à une action dont la cause comprend, entre autres, un élément sensoriel. La plupart du temps, mais pas toujours, cet élément sensoriel est d'ordre visuel. Ainsi, dans une action sensorimotrice, au sens où on comprend habituellement ce type d'action, l'agent perçoit d'abord une situation environnementale, et le mouvement sélectionné par l'agent est fonction de cette perception. Par exemple, le golfeur perçoit le trou

[1] Cependant, comme on va le voir, je me démarque de Davidson en ce qui concerne la cause interne de l'action. Je ne crois pas, contrairement à Davidson, que la cause psychologique proximale de l'action soit constituée d'une croyance et d'un désir.

où il veut envoyer sa balle et le mouvement exécuté par ce golfeur est fonction de cette perception. Bien qu'on puisse parler en ce sens des actions sensorimotrices, j'emploierai l'expression « action sensorimotrice » dans un sens plus restreint. Lorsqu'il y a une action sensorimotrice, il y a certes, comme nous l'avons vu, une dépendance de l'action vis-à-vis la perception, mais il y a aussi une dépendance de la perception vis-à-vis l'action. Ainsi, non seulement ce que l'agent perçoit cause une action chez lui, mais ce que l'agent exécute comme mouvement a aussi un effet causal sur sa perception. Il y a donc, au sens où je comprends l'expression « action sensorimotrice », un couplage ou feedback entre la perception et l'action. À ce titre, nous pouvons voir que l'action du golfeur n'est pas sensorimotrice. Certes, l'action du golfeur change son environnement visuel, mais la perception de ce changement environnemental n'agit pas en retour sur l'action. Autrement dit, dans l'action du golfeur, il n'y a pas de feedback et cela suffit pour écarter cette action des actions de type sensorimoteur.

Il existe au moins deux types d'action sensorimotrice. Comme je viens de le dire, l'action sensorimotrice se caractérise par un couplage sensorimoteur. On retrouve un couplage sensorimoteur à chaque fois que les outputs moteurs (les mouvements) et les inputs sensibles (les états de l'environnement) sont en relation de rétroaction mutuelle (feedback). Le couplage sensorimoteur (la rétroaction mutuelle entre outputs et inputs) peut être de deux types, d'où les deux types d'action sensorimotrice. Le couplage sensorimoteur peut être « serré » ou non. Dans une partie d'échecs, le coup de mon adversaire (l'input) cause chez moi une réponse (l'output ou mouvement), et ma réponse cause un nouveau coup chez mon adversaire, lequel coup va causer une autre réponse chez moi, etc. On retrouve donc ici un couplage sensorimoteur. Ce couplage sensorimoteur, cependant, n'est pas « serré ». Mes coups, de même que ceux de mon adversaire, ne requièrent pratiquement aucun « timing ». Lorsque mon adversaire joue son coup, je peux jouer mon coup immédiatement ou, au grand déplaisir de mon adversaire, attendre le temps qu'il me plaira. Donc, l'état de l'environnement (le coup de mon adversaire) ne contraint pas temporellement mon propre coup. C'est ce que je veux dire lorsque je dis que ce couplage sensorimoteur ne requiert presque aucun « timing ». Et c'est à ce type de couplage sensorimoteur que je pense lorsque je parle d'un couplage sensorimoteur qui n'est pas « serré ».

Un couplage sensorimoteur « serré » est un couplage où l'occurrence de l'input contraint temporellement l'output. Dans ce type de couplage sensorimoteur, il y a du « timing ». Les actions sensorimotrices ayant un couplage sensorimoteur « serré » sont très nombreuses. Par exemple, lorsque je suis en train de jouer au tennis, le coup de mon adversaire contraint temporellement mon propre coup : si j'attends trop longtemps avant de frapper la balle, je ne serai pas en mesure de jouer mon coup.

On retrouve donc deux types d'action sensorimotrice : Une action sensorimotrice ayant un couplage sensorimoteur « serré » et une action sensorimotrice ayant un couplage sensorimoteur qui n'est pas « serré ». Ma thèse s'applique principalement à une action sensorimotrice ayant un couplage sensorimoteur « serré ». Dans ma thèse, je prétends donc déterminer ce qui caractérise le passage à l'expertise du joueur de tennis, et non pas ce qui caractérise le passage à l'expertise du joueur d'échecs.

1.1. Un modèle causal de l'action

Il s'agit maintenant de déterminer les causes internes de l'action et de déterminer la façon dont ces causes s'articulent entre elles pour causer l'action.

Davidson considère que la cause interne de l'action est constituée des croyances et des désirs de l'agent. Ce n'est pas la position que je défends au cours de ma thèse. Je ne nie pas que les croyances et les désirs de l'agent interviennent dans la production de l'action, mais je considère que ces deux variables (les croyances et les désirs) sont des causes distales de l'action. Je déterminerai plus tard (chapitre 2) le rôle des croyances et des désirs dans la production de l'action, mais, pour l'instant, je suis intéressé à déterminer les causes proximales de l'action.

Mon modèle de l'action fait intervenir trois causes proximales de l'action. Premièrement, pour agir, disons faire X, j'ai besoin d'une intention, de l'intention de faire X. Je reviendrai plus tard sur ce que j'entends par la notion d'intention. Pour l'instant, on peut se contenter de notre notion vague et spontanée d'intention. Mon modèle fait aussi intervenir comme cause de l'action une perception d'affordance et une perception de variable optique de contrôle. Je

vais maintenant donner une définition préliminaire des notions de perception d'affordance et de perception de variable optique de contrôle[2].

Les affordances, comme on le sait depuis Gibson (1986), sont des propriétés des objets de l'environnement[3]. L'affordance d'un objet se caractérise comme une possibilité d'action; l'affordance d'un objet spécifie ce qui peut être fait avec cet objet. Ainsi, la chaise « afforde » l'action de s'asseoir, si bien que, lorsqu'il perçoit cette affordance, l'agent voit qu'il peut s'asseoir. De même, lorsque l'agent voit une ouverture dans le mur (l'ouverture « afforde » l'accès), il voit qu'il peut s'y engouffrer.

Mais la perception d'une possibilité d'action ne permet pas de contrôler le mouvement. Pour qu'un élément puisse contrôler le mouvement, il faut que cet élément spécifie le succès ou l'erreur du mouvement en cours d'exécution. C'est ainsi que, sur la base de l'information fournie par cet élément, le système de contrôle peut corriger (ou maintenir) son action. Or, l'affordance d'un objet ne présente rien d'autre qu'une possibilité d'action; elle ne dit pas si l'action en cours d'exécution est un succès ou une erreur. La perception de l'affordance ne peut donc pas servir au contrôle de l'action. Il faut donc trouver un élément qui puisse contrôler le mouvement.

Une variable optique de contrôle (la troisième cause de l'action) est un pattern du flux optique[4]. La variable optique de contrôle accompagne l'exécution du mouvement et elle spécifie si le mouvement en cours d'exécution est un succès ou un échec. La variable optique de contrôle peut donc contrôler le mouvement. Ainsi, si cette variable s'écarte un tant soit peu de la valeur souhaitée (la valeur qui spécifie le succès du mouvement en cours d'exécution), l'agent change son mouvement de manière à ce que la variable optique retrouve la valeur souhaitée[5,6].

---

[2] Le chapitre 3 se penche plus particulièrement sur la notion d'affordance et de perception d'affordance, et le chapitre 6 se penche sur la notion de variable optique de contrôle.

[3] Comme on va le voir au chapitre 3, les affordances sont des propriétés environnementales particulières. Il s'agit de propriétés relationnelles. Je n'en dis pas plus pour l'instant.

[4] Le flux optique correspond aux changements de la structure optique qui se produisent lorsque l'observateur se déplace ou lorsque l'environnement se déplace par rapport à l'observateur. Les changements de la structure optique peuvent être représentés par un champ vectoriel à deux dimensions, où chaque vecteur représente la vitesse à laquelle un élément de la structure optique se déplace par rapport à l'observateur.

[5] Donc, la variable optique est fonction des mouvements de l'agent.

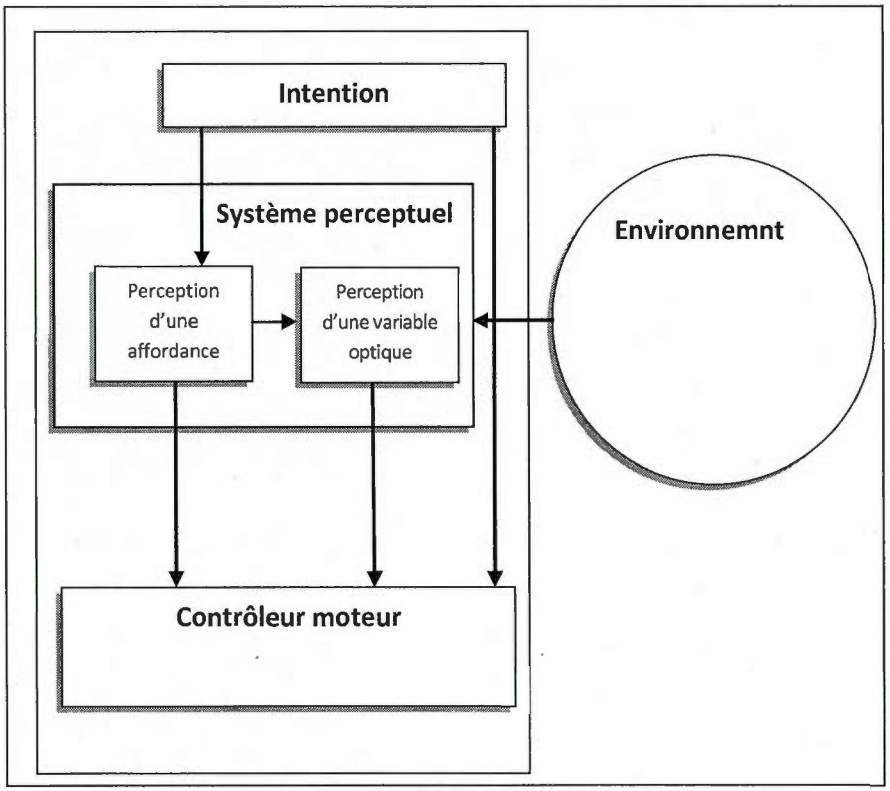


Figure 1.1 Les relations causales entre les différents éléments qui déterminent l'action.

La figure 1.1 met en scène les trois causes dont il vient d'être question. Comme on le voit, ces trois causes produisent un mouvement en agissant sur le contrôleur moteur. Il faudrait dire quelques mots sur le contrôleur moteur. En définissant le contrôleur moteur, nous pourrons voir comment ces trois causes s'articulent entre elles pour produire une action. Aussi, on retrouve dans la figure 1.1 deux relations causales particulières, une qui va de l'intention vers la perception de l'affordance et une autre qui va de la perception de l'affordance vers la perception de la variable optique de contrôle. Commençons par définir ces deux relations causales.

---

[6] Mon modèle causal de l'action fait intervenir une variable de contrôle qui est de type visuel ou optique. Cependant, il est vraisemblable de penser que les variables sensorielles de type optique n'ont pas le monopole du contrôle. Les autres dimensions sensorielles (proprioceptive, kinesthésique, etc.) ont certainement aussi leur variable de contrôle. Donc, même si, au cours de ce texte, je ne parle que de variables de contrôles de type optique, mes propos devraient pouvoir s'étendre aux autres dimensions sensorielles.

L'intention cause (conjointement avec l'environnement) la perception de l'affordance. À première vue, cette action causale de l'intention sur le système perceptuel peut paraître superflue. L'environnement agit déjà causalement sur le système perceptuel. Dans ces circonstances, pourquoi, en plus de l'action causale de l'environnement sur le système perceptuel, avons-nous besoin d'introduire une relation causale qui part de l'intention et qui va vers le système perceptuel?

L'environnement contient une multitude d'affordances. Lorsque l'agent perçoit son environnement, il perçoit donc une multitude d'affordances. Or, l'agent ne peut pas percevoir toutes ces affordances; à tout le moins, il ne peut pas faire attention à un grand nombre de ces affordances. L'argument qui appuie cette dernière proposition est le suivant. Selon notre modèle de l'action, la perception d'une affordance a une force causale sur l'action; elle pousse l'agent à agir. Donc, si l'agent perçoit ou fait attention à une multitude d'affordances en même temps, il sera porté à exécuter une multitude d'actions en même temps[7]. Or, cela est inadmissible. Il faut donc faire intervenir un élément qui agit sur le système perceptuel et qui fait en sorte que l'agent perçoit une seule affordance, celle qui réfère à une action qui satisfait l'intention. C'est ce que l'intention permet de faire en agissant causalement sur le système perceptuel. Donc, ici, la fonction de l'intention est de diriger l'attention du système perceptuel vers une affordance plutôt que vers les autres (vers l'affordance qui satisfait l'intention).

Cette thèse, que l'intention dirige l'attention de l'agent vers des contenus perceptuels plutôt que d'autres, possède une plausibilité psychologique. Par exemple, si je suis dans un stade où se joue un match de tennis, je ferai attention à des choses complètement différentes selon que j'ai l'intention de regarder le match ou selon que j'ai l'intention d'examiner l'architecture du stade.

La figure 1.1 nous dit également que la perception de l'affordance agit causalement sur la perception de la variable optique. L'action causale de la perception d'affordance sur la perception de la variable optique est analogue à l'action causale de l'intention sur la perception de l'affordance : elle consiste à diriger l'attention. Ainsi, idéalement (c.-à-d. au

---

[7] Comme on va le voir plus tard, cela est vrai uniquement à condition que l'intention qui agit sur le système moteur soit une intention très générale.

terme de l'apprentissage), la perception de l'affordance dirige l'attention du système perceptuel vers la bonne variable optique, c'est-à-dire la variable optique qui permet de réaliser l'affordance. Cette thèse, qu'un contenu perceptuel (en l'occurrence, la perception d'une affordance) dirige l'attention vers un autre contenu perceptuel (en l'occurrence, la perception d'une variable optique), est aussi une thèse qui m'apparaît plausible. Voici un exemple qui appuie cette thèse. Si je me lève le matin et que je me dirige vers la cuisine, je peux percevoir la cafetière. Cette perception peut très bien en causer une autre, disons la perception qui porte sur la quantité de café restant dans la cafetière.

Il est maintenant temps de voir comment les trois causes susmentionnées – intention, perception d'affordance et perception de variable optique de contrôle – s'articulent entre elles pour causer le mouvement. Ces trois causes produisent une action en agissant sur le contrôleur moteur. Il reste donc une dernière chose à faire avant de voir comment ces trois causes travaillent conjointement pour produire une action : il reste à définir le contrôleur moteur.

Le contrôleur moteur est un réseau de neurones. Sa fonction est d'envoyer des commandes motrices aux muscles. Autrement dit, le contrôleur moteur active les muscles. Tel que je le comprends, le contrôleur moteur a ceci de particulier qu'il active les muscles de manière à ce que ceux-ci forment une *synergie* ou *structure coordinative*. Une synergie ou structure coordinative est un groupe de muscles où l'activation de chaque muscle est dans un rapport (spatial et temporel) bien déterminé avec l'activation des autres muscles. Une synergie, ou rapport d'activations déterminé entre les muscles, a ceci de particulier qu'elle permet de réaliser une fonction motrice (par ex. : prendre un objet, lancer un objet, attraper un objet, etc.). Chaque synergie permet donc de réaliser une fonction motrice donnée, et, dans mon modèle, chaque synergie est rattachée à son propre contrôleur moteur. Mon modèle du contrôle est donc modulaire.

Maintenant que le contrôleur moteur et la synergie qu'il cause ont été décrits, voyons comment les trois causes de l'action agissent sur le contrôleur moteur afin de produire un mouvement.

Premièrement, l'intention vient prépotentialiser le (ou les) contrôleur(s) moteur(s). Plus précisément, l'intention vient prépotentialiser tous les contrôleurs moteurs dont l'activation

est nécessaire pour qu'elle (l'intention) soit satisfaite. Par exemple, s'il s'agit d'une intention élémentaire, c'est-à-dire d'une intention qui est satisfaite par la réalisation d'une fonction motrice élémentaire (par ex. : prendre un objet), l'intention va prépotentialiser un seul contrôleur moteur, celui qui cause une synergie dont la fonction motrice correspond à l'action de prendre un objet. Si, au contraire, il s'agit d'une intention complexe, c'est-à-dire une intention qui est réalisée à travers plusieurs actions élémentaires, alors l'intention prépotentialise tous les contrôleurs moteurs qui doivent être mis à contribution pour réaliser l'intention complexe. Par exemple, si je suis un voltigeur au baseball et que le frappeur est au marbre, l'intention d'attraper la balle va venir prépotentialiser au moins deux contrôleurs moteurs : le contrôleur moteur correspondant à la synergie qui cause l'action de courir et le contrôleur moteur correspondant à la synergie qui met en branle les bras du voltigeur et qui réalise la fonction d'attraper la balle.

L'intention contribue donc à la production de l'action en prépotentialisant un (ou des) contrôleur(s) moteur(s). Ensuite, la perception de l'affordance intervient. Comme le montre la figure 1.1, cette perception est fonction de l'intention; autrement dit, l'affordance à laquelle l'agent fait attention est celle qui satisfait l'intention, c'est-à-dire celle qui réfère à un mouvement qui satisfait l'intention. La perception de l'affordance vient aussi agir sur le(s) contrôleur(s) moteur(s). S'il s'agit d'une perception qui porte sur une affordance « élémentaire », c'est-à-dire une affordance qui réfère à un mouvement élémentaire, alors cette perception vient agir sur un seul contrôleur moteur, celui qui permet de satisfaire la perception de l'affordance. S'il s'agit d'une perception qui porte sur une affordance « complexe », alors la perception va venir agir sur tous les contrôleurs moteurs qui doivent être mis à contribution pour réaliser le mouvement complexe auquel réfère l'affordance. Ainsi, si l'on suppose que la balle a été cognée par le frappeur et que cette balle se présente au voltigeur avec l'affordance « il est possible de m'attraper », la perception de cette affordance par le voltigeur va venir agir sur tous les contrôleurs moteurs qui sont impliqués dans l'action d'attraper la balle. On peut donc supposer que les deux contrôleurs moteurs mentionnés ci-dessus (courir et tendre les bras pour attraper) recevront cette perception d'affordance comme input. Et, comme ces deux contrôleurs moteurs ont déjà été prépotentialisés par l'intention, l'action de la perception d'affordance sur ces contrôleurs moteurs est suffisante pour les activer. Donc, sous l'effet de la perception d'affordance, ces

deux contrôleurs moteurs produisent leur synergie respective. À ce stade, donc, un mouvement est initié. Cependant, ce mouvement complexe est imparfait. Comme nous l'avons dit plus haut, la perception de l'affordance ne contrôle pas le mouvement. Autrement dit, nos deux contrôleurs moteurs ne peuvent pas se fier à l'information donnée par la perception de l'affordance pour activer adéquatement leur synergie à chaque moment de l'action. Cette information, comme on l'a dit, est donnée par la variable optique de contrôle.

Chaque contrôleur moteur possède sa variable optique de contrôle[8]. La variable optique de contrôle d'un contrôleur moteur spécifie l'erreur du mouvement que le contrôleur moteur produit grâce à l'activation d'une synergie. Le contrôleur moteur peut donc se fier à la valeur de cette variable pour contrôler le mouvement qu'il produit; il lui suffit d'éliminer à chaque instant l'erreur spécifiée par la variable optique de contrôle. Ainsi, sur la base de l'erreur spécifiée par les variables de contrôle qui servent d'input aux deux contrôleurs moteurs qui sont mobilisés par le voltigeur, ce dernier pourra arriver à temps en dessous de la balle et attraper la balle.

On pourrait émettre certains doutes à l'endroit de notre modèle causal de l'action en soulignant que le rôle des intentions y apparaît complètement accessoire et qu'un modèle plus simple, et tout aussi efficace, pourrait être élaboré. Dans ce modèle plus simple et tout aussi efficace, les intentions ne seraient pas comprises comme causes directes ou proximales de l'action. Il m'apparaît important d'inclure les intentions parmi les causes directes de l'action car, contrairement aux perceptions d'affordances et de variables optiques, les intentions ont un caractère conatif. Pour se convaincre du fait que la perception d'une affordance et d'une variable optique n'ont aucun caractère conatif, il suffit de constater que je peux percevoir une affordance et une variable optique et ne pas me sentir poussé à agir. Par exemple, si je suis un batteur au baseball, la balle peut m'« afforder » l'action « frappe-moi » et je peux percevoir la variable optique qui spécifie un coup de batte qui soit un succès, cela sans être pour autant poussé à frapper la balle. Pour qu'il y ait mouvement, il faut donc faire intervenir un élément qui possède un caractère conatif. C'est l'intention qui fait ce travail. Donc, si, en plus de

[8] Il n'est pas exclu qu'une même variable optique puisse s'appliquer à plus d'un contrôleur moteur. Par exemple, la variable tau (τ), qui spécifie l'imminence du contact, pourrait s'appliquer à plusieurs contrôleurs moteurs.

percevoir l'affordance « frappe-moi » et la variable optique de contrôle, le joueur de baseball a l'intention de frapper la balle, le joueur exécutera le mouvement.

Avant de passer à la prochaine étape, il faudrait dire de manière plus précise ce que j'entends par le terme « intention ».

Une intention est une représentation conceptuelle d'un mouvement. L'intention représente aussi l'état environnemental qui fait suite à ce mouvement. Ainsi, si j'ai l'intention d'attraper le verre en face de moi, je me représente conceptuellement le mouvement qui consiste à tendre le bras vers le verre et je me représente aussi l'état environnemental qui fait suite à ce mouvement, c'est-à-dire ma main qui empoigne le verre[9].

Il faut encore établir quelque chose d'important pour la suite des choses : les intentions de l'agent s'organisent de manière hiérarchique (Bratman, 1987). Autrement dit, comme les différentes intentions forment les différents moments du *plan* de l'agent (Bratman, 1987), le plan de l'agent a une structure hiérarchique. Au sommet de la hiérarchie, on retrouve une intention très générale, qui spécifie un but très général ou très éloigné, et cette intention est réalisée à travers quelques intentions plus particulières, qui spécifient des buts qui sont des moyens relativement à l'intention de plus haut niveau. Et ces dernières intentions sont elles-mêmes réalisées à travers des intentions encore plus particulières. Cette division des intentions en des intentions encore plus particulières prend fin lorsque le niveau des habiletés et des habitudes est atteint; à ce niveau, il n'est plus nécessaire de se construire un plan pour diriger son action. En fait, le niveau des habiletés et des habitudes est pris en charge par les perceptions d'affordances et les perceptions de variables optiques de contrôle. Ce sont les intentions du niveau inférieur de la hiérarchie qui contrôlent le système perceptuel et, donc, la perception des affordances.

Le plan hiérarchique n'a pas à être établi tout d'un coup, du sommet de la hiérarchie jusqu'à sa base (Bratman, 1987). L'ajout de nouvelles intentions, plus particulières, peut attendre le moment où l'agent est confronté à son nouvel environnement. Par exemple, si j'ai l'intention de faire mon marché, les intentions qui vont réaliser cette intention de haut niveau ne peuvent se présenter que lorsque je serai rendu au marché (le nouvel environnement). À ce moment, lorsque je serai confronté à ce nouvel environnement, je pourrai spécifier mon

[9] Cette définition des intentions n'est pas complète. Le chapitre 9 va la compléter.

intention d'aller au marché par des intentions plus particulières : je pourrai avoir l'intention de tourner dans cette allée du marché et avoir l'intention de prendre ensuite cette autre allée.

1.2 La thèse que je soutiens

Mon modèle causal de l'action fait intervenir quatre éléments : des intentions, des perceptions d'affordances, des perceptions de variables optiques de contrôle ainsi que des contrôleurs moteurs. Je soutiens que le passage à l'expertise se produit à travers des changements apportés à ces quatre éléments. Voici ce en quoi consistent ces changements.

Premier changement : le changement moteur. Reprenons l'exemple d'un agent qui veut se diriger vers un endroit (son but) et qui, en percevant cet endroit, perçoit l'accessibilité, c'est-à-dire l'affordance « il est possible de s'approcher ». Comme on l'a dit plus haut, le mouvement est contrôlé par une variable optique de contrôle. Supposons que la perception de cette affordance fasse en sorte que l'agent perçoive la *bonne* variable optique de contrôle; il perçoit le focus d'expansion du flux optique[10]. Mon point, maintenant, est le suivant : même en percevant cette variable optique de contrôle, l'agent peut être incapable de se diriger vers son but. C'est ce qui arrivera si le contrôleur moteur (un réseau de neurones) n'est pas paramétrisé de manière à pouvoir éliminer l'erreur spécifiée par la variable optique de contrôle. Donc, le premier changement consiste à modifier la valeur des paramètres du contrôleur moteur (c.-à-d., plus précisément, à changer le poids des connexions du réseau de neurones) de manière à ce que ce dernier puisse éliminer l'erreur spécifiée par la variable optique de contrôle.

Deuxième changement : le changement perceptuel. Initialement, c'est-à-dire lorsqu'il est novice, l'agent ne perçoit pas la bonne affordance (celle qui satisfait l'intention) et la bonne variable optique de contrôle (celle qui satisfait[11] la perception de l'affordance), tandis que,

---

[10] Le focus d'expansion correspond au point central à partir duquel les éléments de la structure optique prennent de l'expansion lorsque l'agent se déplace vers l'avant. À partir de la valeur que prend le focus d'expansion, il est possible de dire si l'action consistant à se déplacer vers un objet est un succès ou une erreur. Le focus d'expansion permet donc de contrôler l'action qui consiste à se déplacer vers un objet.

[11] Quand je dis qu'une variable optique *satisfait* une perception d'affordance, je veux essentiellement dire que cette variable optique de contrôle permet de réaliser un mouvement qui correspond au mouvement auquel réfère la perception de l'affordance.

lorsqu'il est expert, il perçoit la bonne affordance et la bonne variable optique de contrôle. Le deuxième changement est celui qui nous fait passer d'une perception d'affordance et de variable optique inappropriée à une perception d'affordance et de variable optique appropriée.

Nous retrouvons donc deux types de changements perceptuels : le premier type porte sur la variable optique; le second, sur l'affordance. Est-ce que, du point de vue de la psychologie, il est sensé de dire que l'agent doit passer par ces deux changements perceptuels avant d'acquérir une expertise motrice? Plusieurs expériences ont montré que, pour acquérir une nouvelle habileté, l'agent devait apprendre à percevoir la bonne variable optique de contrôle (Fajen *et al.*, 2009), et, que l'agent ait besoin d'une période d'apprentissage pour percevoir l'affordance qui va lui permettre de réaliser le mouvement qui satisfait son intention, cela aussi est une thèse extrêmement plausible. Prenons par exemple un joueur de soccer débutant. Celui-ci est devant le but adverse et il a l'intention de mettre le ballon dans le but. Malgré son intention, il tire régulièrement sur le gardien au lieu de tirer dans l'espace vacant, ce qui lui permettrait de compter. Le fait que le joueur tire sur le gardien peut s'expliquer par le fait que, au lieu de regarder l'espace laissé vacant, le joueur regarde le gardien. Le joueur ne perçoit donc pas la bonne affordance; c'est l'espace laissé vacant qui « afforde » l'accès au filet, ce n'est pas le gardien de but. Le même phénomène se produit au tennis. Bien qu'il ait l'intention de frapper la balle dans l'espace laissé vacant, le joueur débutant frappe régulièrement sur le joueur adverse. Encore une fois, on peut expliquer cela par le fait que le joueur ne regarde pas la bonne affordance; au lieu de regarder l'espace laissé vacant, il regarde le joueur adverse.

Donc, en bref, l'acquisition d'une habileté requiert un changement perceptuel; l'agent doit apprendre à percevoir la bonne affordance (celle qui satisfait l'intention) et la bonne variable optique (celle qui satisfait la perception de l'affordance).

Troisième changement : le perfectionnement du plan. Lorsqu'il est novice, l'agent élabore de mauvais plans tandis que, une fois expert, ses plans son maximalement efficaces. La plupart des modifications qui sont apportées au plan touche plus souvent qu'autrement les intentions de bas niveau. Par exemple, si un joueur dispute un match de tennis, ce joueur possède une intention de haut niveau — disons, gagner le match — et cette intention est

réalisée par l'intention de gagner le point, qui est elle-même réalisée par l'intention de jouer un amorti. Et cette dernière intention peut elle-même être réalisé par l'intention de jouer un amorti de ce côté du terrain. Il est fort probable que, au sein de ce plan hiérarchique, les intentions qui seront remplacées par d'autres intentions, de meilleures intentions, soient les intentions du niveau inférieur. Autrement dit, il est probable qu'une amélioration de la performance requiert l'intention de jouer un amorti de l'autre côté du terrain ou qu'il faille changer l'intention de jouer un amorti par l'intention de jouer un coup en fond de terrain. Les intentions de haut niveau – l'intention de gagner le match et l'intention de gagner le point – sont plus rarement révisables. Donc, lorsqu'il s'améliore, l'agent commence par perfectionner ses intentions de bas niveau. Par ailleurs, il est important de noter que ce perfectionnement du plan s'effectue aussi dans la dimension horizontale; autrement dit, l'agent planifie à plus long terme. Ainsi, lorsqu'il devient meilleur, le joueur peut avoir l'intention d'envoyer la balle du côté droit du terrain dans le but de déporter son adversaire de ce côté du terrain, ce qui lui donnera ensuite l'opportunité de jouer la balle de l'autre côté du terrain (du côté gauche). Dans ce cas de figure, le joueur a un plan qui s'étend dans la dimension horizontale : il a l'intention de jouer à droite, ce qui va déporter son adversaire vers la droite, et il a l'intention de jouer à gauche par la suite.

Quatrième changement : l'élimination progressive du plan. Lorsque les plans de l'agent sont maximalement efficaces, que l'agent parvient à percevoir les affordances qui sont appropriées aux intentions de bas niveau, qu'il perçoit les variables optiques de contrôle qui satisfont les perceptions d'affordances et, enfin, qu'il est parvenu à donner une pondération adéquate au réseau de neurones (le contrôleur moteur), il se produit un changement remarquable. En fait, je considère que ce changement constitue l'apport le plus original de ma thèse. Ce changement consiste en ceci : lors de l'acquisition d'une expertise, les intentions de bas niveau son éliminées et le contrôle du système perceptuel (plus précisément, le contrôle des perceptions d'affordances) est pris en charge par les intentions qui sont au niveau immédiatement supérieurs. Et, à mesure que l'agent devient expert, ces intentions immédiatement supérieures sont éliminées au profit d'intentions situées encore plus haut dans la hiérarchie; ce sont alors ces dernières qui contrôlent le système perceptuel. Selon la complexité du type d'activité en jeu, cette élimination des intentions au profit d'intentions supérieures peut continuer jusqu'au moment où l'agent atteint l'intention qui est au sommet

de la hiérarchie. C'est alors cette seule intention qui contrôle l'ensemble de l'activité perceptuelle. Les figures 1.2 et 1.3 comparent un agent débutant et un agent expert.

Donc, lorsqu'il est expert, l'action de l'agent repose sur un plan beaucoup moins complet que lorsqu'il est débutant. Plus précisément, lorsqu'il est expert, l'agent est beaucoup plus réactif; sa réactivité aux affordances et aux variables optiques de contrôle occupe l'essentiel de l'espace du système « cognitif » responsable de l'action. Que l'agent expert se caractérise comme un système réactif, cela a été défendu par plusieurs, dont Brooks (1999).

L'argument de Brooks est relativement simple. L'expert est celui qui est en mesure de gérer les événements inattendus. Gérer les événements inattendus dans un environnement statique ne pose pas vraiment problème. L'agent a alors tout son temps pour résoudre le problème qui lui est posé par l'événement inattendu. En fait, c'est dans un environnement fortement dynamique que la capacité de l'agent à gérer des événements inattendus est mise à l'épreuve. Dans un environnement fortement dynamique, le monde change très rapidement, si bien que le temps pris par l'agent pour gérer l'événement inattendu – c'est-à-dire le temps que prend l'élaboration d'un nouveau plan – peut s'étendre au point de rendre la « solution » (le nouveau plan) de l'agent inappropriée. Dans le chapitre 2, je vais effectivement montrer que, compte tenu de la nature de l'architecture cognitive qui est à la base de nos plans, il y a tout lieu de penser que l'agent n'a pas le temps d'élaborer un nouveau plan lorsqu'un événement inattendu survient au sein d'un environnement dynamique. Donc, comme il est capable de gérer les environnements fortement dynamiques, l'expert doit être, non pas planificateur, mais réactif. Le dernier changement que j'ai introduit ci-dessus, le quatrième changement, est donc justifié.

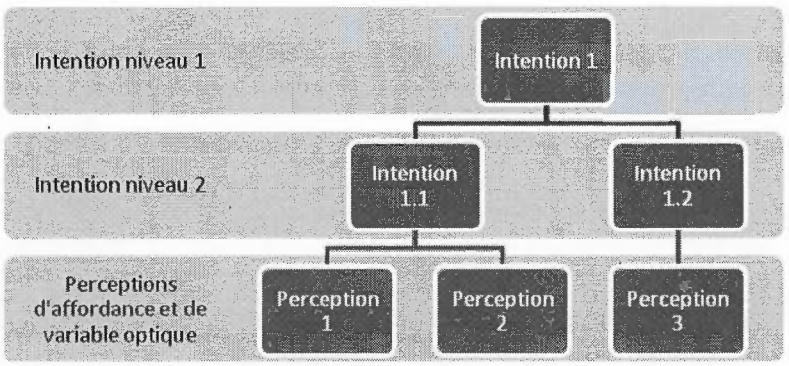


Figure 1.2 Le contrôle des perceptions par les intentions chez le *novice*.

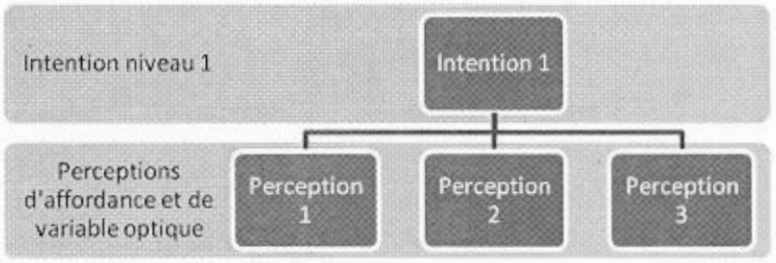


Figure 1.3 Le contrôle des perceptions par les intentions chez l'*expert*.

Donc, ma thèse peut se résumer en ces termes : l'acquisition d'une expertise sensorimotrice (ayant un couplage sensorimoteur « serré ») se caractérise 1) par un changement moteur, 2) par un changement perceptuel qui me permet de percevoir les bonnes affordances et les bonnes variables optiques de contrôle, 3) par l'élaboration de meilleurs plans et, enfin, 4) par l'élimination d'intentions au profit d'intentions toujours plus générales (situées plus haut dans la hiérarchie).

Ma thèse, comme je l'ai dit, consiste à défendre la possibilité de ces quatre changements. Comment vais-je défendre ces quatre types de changements? Chacun de ces changements est le fruit d'un processus d'apprentissage[12]. Donc, si je ne parviens pas à trouver un processus d'apprentissage pour chacun de ces changements, la plausibilité de ma thèse s'en trouvera affaiblie. Donc, pour défendre ma thèse, il faut à tout le moins trouver des processus d'apprentissage pour chacun de ces changements. La défense de ma thèse va consister en grande partie à trouver un processus d'apprentissage pour chacun de ces changements. J'aimerais aussi dire ceci : je m'appliquerai à défendre la cohérence *interne* de ces processus d'apprentissage; je ne m'intéresserai pas à établir leur cohérence *externe*. Autrement dit, je m'appliquerai à montrer que les différents éléments qui composent ces processus

[12] Bien sûr, il se pourrait que ces changements soient le résultat de processus qui ne sont pas des processus d'apprentissage. Par exemple, on pourrait penser ici à un processus de maturation. Cependant, je crois que ces changements proviennent non pas de processus de maturation, mais de processus d'apprentissage. L'argument à l'appui de cette dernière proposition est le suivant. Si un changement est le fruit d'un processus de maturation, alors il a besoin de peu (ou pas) de stimuli pour réaliser ce changement. Or, les quatre changements dont il a été question ont besoin d'un nombre considérable de stimuli pour être réalisés. Donc, ces changements ne proviennent pas d'un processus de maturation.

d'apprentissage sont cohérents entre eux, mais je ne chercherai pas à montrer que ces processus d'apprentissage sont cohérents avec des éléments qui leur sont extérieurs, par exemple avec des faits ou des modèles qui relèvent des neurosciences.

1.3 Division et subdivision de mon texte

Mon texte se divise en deux grandes parties. La première partie porte, non pas sur les changements (et les processus d'apprentissage) qui surviennent lors de l'acquisition d'une habileté, mais sur ce qu'il y a d'invariant chez l'agent, qu'il soit novice ou expert. Comme je l'ai montré lors de ma description du système « cognitif » qui est à la base de l'action du novice et de l'expert, le novice et l'expert basent tous deux leur action sur des intentions, des perceptions d'affordances, des perceptions de variables optiques de contrôle et des contrôleurs moteurs. Donc, dans la première partie, je me pencherai sur les notions d'intention, de perception d'affordance, de variable optique de contrôle et de contrôleur moteur, mais je m'appliquerai surtout à défendre la possibilité de percevoir des affordances. Dans la deuxième partie, je me pencherai sur les changements que j'ai décrits ci-dessus. En fait, comme je l'ai dit, je me pencherai plus précisément sur les processus d'apprentissage qui rendent possibles ces changements. Voici, plus précisément, un avant-goût de ce dont ces chapitres vont parler.

Deuxième chapitre. L'action du novice se caractérise par le fait qu'elle s'appuie sur un plan détaillé. L'action de l'expert, au contraire, s'appuie sur un plan très général, où seules subsistent des intentions situées très haut dans la hiérarchie. Les plans sont le fruit d'une architecture cognitive. Mon défi est de trouver une seule et même architecture cognitive qui vaille tant pour les plans du novice que pour les plans de l'expert. En effet, il serait beaucoup trop lourd théoriquement de poser l'existence de deux architectures cognitives distinctes, une pour les plans du novice et une autre pour les plans de l'expert. J'ai trouvé l'essentiel de cette architecture chez Bratman *et al.* (1988).

Dans les chapitres 3 à 5, je me penche sur la notion d'affordance et, plus précisément, sur la notion de perception d'affordance. Il est important de se pencher sur cette notion, d'une part, car elle est centrale à ma thèse et, d'autre part, car il s'agit d'une notion assez controversée.

Chapitre 3 Dans ce chapitre, je définis la notion d'affordance et la notion de perception d'affordance. La notion d'affordance a été définie conformément à la position de Turvey (1992). Plus précisément, les affordances sont comprises comme des propriétés dispositionnelles. Dans la deuxième partie du chapitre 3, j'aborde la question de la perception des affordances. Conformément à Gibson (1979/1986), la perception des affordances est comprise comme une perception directe ou non inférentielle. L'argument général qui sous-tend la thèse selon laquelle la perception des affordances est une perception directe est le suivant. Percevoir une affordance, c'est posséder l'information sur l'affordance. Or, l'information sur l'affordance est contenue dans la structure optique[13]. Donc, percevoir une affordance, c'est détecter la structure optique. Or, la détection de la structure optique est directe, non inférentielle. Donc, la perception d'une affordance est directe, non inférentielle.

Chapitre 4. L'une des attaques importantes ayant été adressées à la possibilité de la perception directe est celle de Fodor et Pylyshyn. Cette critique est parue dans leur article de 1981, *How direct is visual perception? Some reflections on Gibson's "ecological approach"*. Dans le chapitre 4, je tente de répondre aux critiques de Fodor et Pylyshyn.

Chapitre 5. Dans le chapitre 5, je défends toujours la possibilité de percevoir des affordances. L'une des critiques adressées à la possibilité de percevoir des affordances consiste à dire qu'il n'y a pas de mécanisme concevable qui puisse rendre compte de la perception des affordances. Comme nous l'aurons vu au cours du chapitre 3, la perception des affordances consiste essentiellement à détecter une structure optique qui se caractérise comme un invariant d'invariants. Dans le chapitre 5, je m'emploierai donc à trouver un mécanisme qui permet la détection des invariants d'invariants. Pour trouver ce mécanisme, je me suis inspiré du réseau de neurones semi-récurrent de Elman (1990).

Chapitre 6. Dans le chapitre 6, je définis plus précisément la notion de variable optique de contrôle et de contrôleur moteur. Plus précisément, je donne quelques exemples de variables optiques de contrôle et je montre que le contrôleur moteur est un contrôleur par feedback. Je dégage ensuite les problèmes auquel doivent faire face les contrôleurs par feedback et je tente de solutionner ces problèmes.

---

[13] La structure optique qui spécifie une affordance est un invariant d'invariants.

La deuxième partie couvre les chapitres 7 à 9. Comme je l'ai dit, cette deuxième partie tente de trouver des mécanismes d'apprentissage qui rendent compte des changements qui ont été décrits ci-dessus.

Chapitre 7. Dans le chapitre 7, je tente de trouver un mécanisme d'apprentissage qui puisse rendre compte du changement 1, c'est-à-dire le changement moteur qui nous fait passer d'une exécution motrice inadéquate à une exécution motrice adéquate. Ce mécanisme d'apprentissage a pour fonction de changer la pondération du réseau de neurones de manière à ce que le contrôleur moteur émette des commandes motrices qui vont éliminer l'erreur spécifiée par la variable optique de contrôle (son input). Pour trouver ce mécanisme d'apprentissage, je me suis fortement inspiré du mécanisme d'apprentissage de Ashby, tel que présenté dans *Design for a Brain* (1960).

Chapitre 8. Dans le chapitre 8, je tente de trouver un mécanisme d'apprentissage qui puisse rendre compte du changement 2, c'est-à-dire le changement perceptuel qui nous fait passer d'une mauvaise perception d'affordance (et d'une mauvaise perception de variable optique) à une bonne perception d'affordance (et à une bonne perception de variable optique). Le mécanisme d'apprentissage que j'ai trouvé est un apprentissage par renforcement. Plus précisément, il s'agit de l'apprentissage par renforcement tel que développé par Sutton et Barto (1998).

Chapitre 9. Dans ce chapitre, je tente de trouver un mécanisme d'apprentissage qui puisse nous expliquer le changement 4, c'est-à-dire le changement qui nous fait passer d'un agent qui contrôle les perceptions d'affordances à l'aide d'un plan très détaillé à un agent qui contrôle le système perceptuel à l'aide d'un plan très général (c'est-à-dire un plan où seules subsistent des intentions situées très haut dans la hiérarchie). En apparence, ce changement est un changement du plan. Certes, le plan change, mais, d'un point de vue computationnel, le changement significatif concerne le système perceptuel. L'argument qui me permet d'appuyer cette dernière proposition est le suivant. Selon notre modèle, lorsqu'il est novice, le système perceptuel reçoit plusieurs inputs venant des intentions de bas niveau. Chacun de ces inputs occasionne chez l'agent une perception d'affordance. Or, une fois qu'il est expert, les *mêmes* perceptions d'affordances (les outputs du système perceptuel) sont causées par beaucoup moins d'inputs, puisque les inputs du système perceptuel sont les intentions de

l'agent et que, une fois expert, l'agent possède beaucoup moins d'intentions. Donc, compte tenu que, malgré cette différence au niveau des inputs, le système perceptuel émet toujours les *mêmes* outputs (les mêmes perceptions d'affordances), la capacité computationnelle du système perceptuel à gérer ses outputs (les perceptions d'affordances) *change* considérablement lorsque celui-ci passe de novice à expert. Le fait que l'élimination progressive du plan implique pour l'essentiel un changement qui concerne le système perceptuel suggère ceci : l'apprentissage qui valait pour le changement 2 (le changement perceptuel) devrait aussi valoir pour le changement qui correspond à l'élimination progressive du plan (le changement 4). Comme je l'ai dit, l'apprentissage qui est à la base du changement perceptuel (le changement 2) est un apprentissage par renforcement, et je compte précisément montrer que l'apprentissage par renforcement est à la base du changement qui nous fait passer d'un agent où le système perceptuel est contrôlé par un plan détaillé à un agent où le système perceptuel est contrôlé par un plan très général.

Le lecteur aura certainement noté que je n'ai pas mentionné le changement 3, celui qui se caractérise par une amélioration du plan. J'aurai l'occasion de parler du processus d'apprentissage qui est à la base de ce changement, mais je ne lui consacre aucun chapitre car je serai assez expéditif au sujet de ce processus d'apprentissage. Je serai assez expéditif à propos de ce processus d'apprentissage car je m'intéresse tout particulièrement aux processus d'apprentissage qui font intervenir de manière importante les perceptions d'affordances et de variables optiques de contrôle, et l'apprentissage qui est à la base de l'amélioration du plan fait très peu intervenir ces perceptions.

CHAPITRE II

L'ARCHITECTURE COGNITIVE À LA BASE DU PLAN D'ACTION

L'objectif du chapitre 2 est d'établir l'architecture cognitive à partir de laquelle le plan hiérarchique est élaboré. Il est important d'établir cette architecture cognitive car, à défaut de cette architecture, il n'est pas possible de comprendre la possibilité d'avoir un plan d'action, et notre modèle de l'action, tant pour l'expert que pour le novice, repose en partie sur l'existence d'un plan d'action.

Dans notre architecture, l'intention n'est pas réductible à une croyance et à un désir, comme le soutiennent certains philosophes (Audi (1973); Beardsley (1978); Churchland (1970); Davis (1984)). L'architecture cognitive dont il sera question ne délaisse pourtant pas les croyances et les désirs; elle intègre plutôt croyances, désirs et intentions. Pour employer l'expression consacrée, il s'agit d'une architecture CDI (croyances/désirs/intentions). Bratman (1987) et Bratman *et al.* (1988) sont probablement les pionniers en ce qui concerne les architectures CDI. Ce chapitre s'inspire en premier lieu de ces auteurs.

Mais avant d'aller de l'avant et de décrire cette architecture, il convient de se persuader du bien-fondé de la thèse voulant que les intentions ne soient pas réductibles à un complexe de croyances et de désirs. À cet effet, je vais procéder en deux temps. Je vais d'abord décrire les caractéristiques qui sont attribuées aux intentions, et je vais ensuite montrer que ces caractéristiques ne sont pas explicables en termes de croyances et de désirs. Ce n'est qu'à la suite de ces deux étapes, c'est-à-dire lorsque j'aurai montré que les intentions sont des états mentaux à part entière, que je présenterai l'architecture CDI.

2.1 Les intentions comme états mentaux à part entière

Avant de déterminer le caractère non réductible des intentions, j'aimerais me pencher sur ce qui fait qu'un état mental est un état mental. Suivant Fodor (1983), un état mental est composé d'une attitude et d'une représentation mentale. Donc, en guise de considérations préliminaires, j'aimerais me pencher sur les notions de représentation et d'attitude.

2.1.1. Les représentations et les attitudes

Les représentations indiquent quelque chose. Ainsi, toute croyance renvoie à un état de choses possible ou imaginable. Il en va de même pour nos intentions. Les intentions renvoient aux buts de l'agent. La même chose vaut pour la perception. Toute perception renvoie à un état de choses du monde. La capacité de la représentation à indiquer quelque chose peut provenir de plusieurs facteurs. Elle peut provenir de la ressemblance entre la représentation et l'état de choses représenté. Elle peut aussi provenir du fait que la représentation a une situation bien précise au sein du réseau fonctionnel (Haugeland, 1985). C'est là l'explication fonctionnaliste. Il existe probablement d'autres façons d'indiquer l'état de choses représenté, mais je vais adopter la perspective informationnelle (Dretske, 1981, 1988; Cummins, 1989). Selon la perspective informationnelle, un état physique indique (ou informe sur) autre chose si, et seulement si, l'état physique en question est corrélé nomologiquement à l'état de choses. Par exemple, ce qui nous fait dire que l'orientation du tournesol indique (ou nous informe sur) la position du soleil, c'est que l'orientation du tournesol est corrélée nomologiquement à la position du soleil. Donc, nous dirons d'un état neuronal qu'il indique (ou donne de l'information sur) un état de choses à condition que l'état neuronal soit corrélé nomologiquement à cet état de choses.

Mais il faut voir que, pour être une représentation, il ne suffit pas d'indiquer un état de choses; il ne suffit pas d'être corrélé nomologiquement à l'état de choses. Pour que la structure indicative (l'état neuronal qui indique l'état de choses) soit une représentation, la structure indicative doit aussi avoir la *fonction* d'indiquer ce qu'elle indique (Dretske, 1988). Essayons de voir pourquoi il en va ainsi.

Une structure indicative indique une multitude de choses. En fait, comme on l'a vu, une structure indicative indique toutes les propriétés avec lesquelles elle est corrélée nomologiquement, et elle est corrélée nomologiquement avec une multitude de propriétés. Or, la structure indicative ne représente pas toutes ces choses. Par exemple, le baromètre indique la pression atmosphérique ainsi que l'altitude, mais il ne représente qu'une seule chose, soit la pression atmosphérique. D'où vient que, bien qu'il indique aussi l'altitude, le baromètre représente seulement la pression atmosphérique? Cela vient du fait que le baromètre a reçu la *fonction* d'indiquer la pression atmosphérique et qu'il n'a pas reçu la fonction d'indiquer l'altitude. Ainsi, pour avoir une représentation, il n'est pas suffisant d'avoir une structure qui indique une propriété; il faut aussi que cette structure ait reçu la fonction d'indiquer cette propriété. Donc, un état neuronal représente un état de choses donné parce qu'il a reçu la fonction d'indiquer cet état de choses. Ensuite, la question est de déterminer le facteur qui fait en sorte que cet état neuronal (structure indicative) reçoit la fonction d'indiquer cet état de choses plutôt qu'un autre. Pour ce qui est des objets fabriqués par l'homme, comme le baromètre, ce facteur est le concepteur lui-même. Mais il n'y a pas de concepteur à la base du cerveau et de ces états représentationnels. Suivant Dretske (1988), c'est l'apprentissage qui donne ces fonctions d'indiquer au cerveau. Et selon Millikan (1984), c'est l'évolution.

Comme je l'ai dit plus haut, les états mentaux ne sont pas seulement des représentations, ce sont des attitudes qui ont pour contenu des représentations. Par exemple, la croyance que la neige est blanche et le désir que la neige soit blanche sont deux états mentaux différents ayant la même représentation pour contenu. Ce qui distingue ces deux états mentaux, c'est l'attitude. Dans un cas, il s'agit de l'attitude doxique (relative à la croyance) et, dans l'autre cas, il s'agit de l'attitude volitive (relative au désir). Essayons maintenant de définir la notion d'attitude. Une attitude est individuée par ses rôles fonctionnels (Fodor, 1983; Bratman, 1987). Par exemple, entre autres choses, la croyance est individuée par le fait qu'elle sert d'input au raisonnement; c'est là l'une de ses *fonctions*. Ainsi, les intentions, qui sont des attitudes, doivent aussi être définies par leurs rôles fonctionnels. Bratman (1987) attribue quatre caractéristiques ou rôles fonctionnels aux intentions. Par conséquent, si les intentions sont réductibles à un complexe de croyances et de désirs, ces quatre rôles fonctionnels doivent être explicables par les fonctions qui sont propres aux désirs et aux croyances.

Maintenant, comme je l'ai annoncé ci-dessus, mon but est de montrer qu'une telle explication par les désirs et par les croyances n'est pas possible. Si je parviens à montrer cela, j'aurai montré que les intentions ne sont pas réductibles à un complexe de croyances et de désirs.

2.1.2. La non réductibilité des intentions

Bratman (1987) attribue quatre caractéristiques aux intentions. Tout d'abord, Bratman considère que, tout comme les désirs, les intentions sont des pro-attitudes : lorsqu'elles sont accompagnées de la croyance appropriée, elles *motivent* l'agent à réaliser une action. Ainsi, si j'ai l'intention (ou le désir) de faire plaisir à ma compagne et que je crois qu'une rose lui ferait plaisir, je serai porté à lui offrir une rose. Cependant, l'intention est une pro-attitude qui se distingue du désir. Je peux avoir le désir de faire plaisir à ma compagne et *ne pas* lui offrir cette rose, même si rien d'imprévu ne survient. Je pourrais en effet avoir un désir conflictuel qui consisterait à vouloir économiser mon argent. Donc, les désirs « are merely *potential influencers* of action. » (Bratman, 1987, p. 16) Par contre, si j'ai l'*intention* de faire plaisir à ma compagne et que rien d'imprévu ne survient, je vais lui offrir cette rose. L'intention possède donc une valeur de *contrôle* sur l'action. « Intentions are, whereas ordinary desire are not, conduct *controlling* pro-attitudes. » (Bratman, 1987, p. 16)

Par ailleurs, ce qui permet de nous persuader du fait que les intentions possèdent une valeur de contrôle sur l'action, c'est qu'il est irrationnel d'adopter simultanément des intentions incompatibles ou conflictuelles. Pour comprendre pourquoi l'irrationalité associée à la coexistence d'intentions incompatibles signifie que les intentions possèdent un contrôle sur l'action, il suffit de se demander pourquoi la coexistence d'intentions incompatibles est irrationnelle. Il est irrationnel d'avoir simultanément des intentions incompatibles car il est impossible d'*exécuter* simultanément des actions incompatibles. Donc, la notion d'intention réfère bel et bien au contrôle de l'action.

La deuxième caractéristique des intentions est la suivante : lorsque j'adopte une intention, la délibération à propos de ce qu'il faut faire prend fin. Ainsi, si j'ai pris la décision d'aller à la bibliothèque de l'UQAM, je cesse alors de me demander si je dois ou non aller à la bibliothèque. Les choses sont pour ainsi dire réglées. Bien sûr, je peux reconsidérer mon intention, mais cela ne se produit que si je reçois de nouvelles informations ou si ce que je

veux ou désire change. « Retention of my prior intention and nonreconsideration is, so to speak, the "default option". » (Bratman, 1987, p. 17)

La troisième caractéristique concerne le raisonnement procédant des fins aux moyens (means-end reasoning). Lorsque j'ai une intention, mon intention me dispose à procéder à un raisonnement en vue de trouver des moyens qui vont me permettre de réaliser mon intention (ou ma fin). Ainsi, si j'ai l'intention d'aller à la bibliothèque de l'UQAM, cette intention va servir d'input pour un raisonnement qui va me permettre de trouver des moyens pour réaliser cette intention, et ces moyens sont eux-mêmes des intentions. Ainsi, à l'issue du raisonnement procédant des fins aux moyens, je pourrais adopter l'intention de prendre le bus ou le métro (un moyen pour aller à la bibliothèque de l'UQAM). Ici, je tiens à relativiser l'importance de cette troisième caractéristique. Dans le modèle de l'action que je propose, c'est lorsqu'il est au stade de novice que l'agent élabore un plan détaillé de son action; lorsqu'il est expert, l'agent se contente d'intentions très générales (situées très haut dans la hiérarchie). Donc, si mon modèle de l'action est valable, ce n'est que lorsqu'il est novice que les intentions de l'agent servent d'input à un raisonnement procédant des fins aux moyens.

La quatrième caractéristique concerne encore le raisonnement procédant des fins aux moyens. Les intentions préalables font en sorte que toute intention issue d'un raisonnement procédant des fins aux moyens doit être compatible avec les intentions préalables. Ainsi, si j'ai l'intention préalable de laisser mon auto à ma compagne et que je ne possède qu'une seule auto, je ne peux pas décider (avoir l'intention) d'aller à la bibliothèque de l'UQAM en auto. Ce moyen en vue de réaliser mon intention d'aller à la libraire est incompatible avec mon intention préalable de laisser mon auto à ma compagne et avec la croyance que je ne possède qu'une seule auto. Je ne peux donc pas adopter cette intention-moyen. En bref, la contrainte de compatibilité pose que « [...] it should be possible for her plans, taken together, to be executed successfully in a world in which her beliefs are true. » (Bratman *et al.*, 1988, p. 350)

Je viens de dire que, du fait de l'incompatibilité entre l'intention de laisser son auto à sa compagne et l'intention d'aller à la bibliothèque en auto, l'intention de plus bas niveau (l'intention-moyen) était d'emblée exclue; que l'agent devait exclure l'intention d'aller à la bibliothèque en auto (le moyen permettant de réaliser son intention préalable d'aller à la

bibliothèque). Cela n'est pas toujours le cas. Si l'intention-moyen satisfait certain(s) critère(s) jugé(s) important(s) par l'agent, la présence de cette intention peut remettre en question l'autre intention, celle qui a été préalablement établie (l'intention de laisser mon auto à ma compagne). Lorsque l'intention-moyen satisfait ces critères, un processus de délibération se met en branle; l'agent va délibérer sur les deux intentions incompatibles et il n'est pas impossible qu'il décide de choisir l'intention d'aller à la bibliothèque en auto et que, donc, il décide de ne plus laisser son auto à sa compagne.

Les intentions se caractérisent donc par ces quatre caractéristiques : 1) elles contrôlent nos actions; 2) elles mettent fin à la délibération; 3) elles servent d'input à un raisonnement procédant des fins aux moyens; et 4) elles imposent un « filtre » de compatibilité. Alors si les intentions sont effectivement réductibles à un complexe de croyances et de désirs, il faut que, d'une manière ou d'une autre, nos croyances et nos désirs possèdent ces quatre caractéristiques. Bratman (1987) montre que nos croyances et nos désirs ne possèdent pas ces quatre caractéristiques.

La tentative de réduction la plus évidente consiste à identifier les intentions à des désirs dominants : avoir l'intention de faire X, c'est avoir le désir dominant de faire X. Et avoir un désir dominant de faire X, c'est désirer davantage faire X que désirer faire les actions qui sont incompatibles avec l'action X. Bratman (1987) montre que le désir dominant ne possède pas les quatre caractéristiques qui viennent d'être posées.

Considérons d'abord la seconde caractéristique présentée ci-dessus, celle qui pose que l'intention met fin à la délibération. Demandons-nous si le fait d'avoir un désir dominant met fin à la délibération. Clairement, il faut répondre par la négative à cette interrogation : le désir dominant ne met pas fin à notre délibération. Pour s'en persuader, il suffit de se rappeler la fois où nous avons été le sujet d'un désir dominant et de constater qu'au moment où nous avons été habités par ce désir dominant, la délibération ne prend pas fin. Par exemple, l'alcoolique peut très bien avoir le désir dominant de boire un coup sans toutefois cesser de songer à la possibilité de rester sobre. En effet, il est plausible de penser qu'au moment même où il est habité par le désir dominant de boire, l'esprit de l'alcoolique soit le théâtre d'un conflit furieux entre son désir dominant de boire et son désir de rester sobre. Plus près de nous, on peut penser qu'il arrive souvent qu'un étudiant ait le désir dominant de faire la fête

avec ses amis tandis qu'il ne cesse pas de soupeser la possibilité de rester tranquillement chez lui à faire son travail de fin de session. Donc, clairement, le désir dominant n'arrive pas à expliquer la seconde caractéristique que nous avons attribuée aux intentions; le désir dominant ne met pas fin à la délibération.

Le désir dominant ne permet pas non plus d'expliquer la troisième caractéristique, celle qui pose que le fait d'avoir une intention dispose l'agent à réaliser un raisonnement procédant des fins aux moyens. Le désir dominant, comme on vient de le voir, ne met pas fin à la délibération. Or, si je suis encore en train de délibérer pour savoir si, oui ou non, je dois faire une certaine action, je ne serai certainement pas disposé à chercher des moyens en vue de réaliser cette même action.

Le désir dominant ne permet pas davantage d'expliquer la première caractéristique, selon laquelle l'intention possède un contrôle sur l'action. Le désir dominant, comme on l'a vu, ne règle pas la question de la délibération; il n'y met pas un terme. Donc, au moment où il est temps de passer à l'action, l'agent habité d'un désir dominant peut encore être en train de délibérer pour savoir s'il faut faire ou non cette action. Or, s'il est encore en train de délibérer à ce sujet, l'agent ne sera certainement pas disposé à exécuter cette action.

Enfin, si l'intention est conçue comme un désir dominant, nous ne serons pas en mesure de décrire adéquatement la façon dont l'intention préalable contraint les autres intentions. Me fondant toujours sur Bratman (1987), c'est ce que je vais maintenant essayer de montrer.

Lorsque l'agent est habité par deux intentions incompatibles, il est coupable d'irrationalité. Mais le point important est le suivant : il est *possible* d'être irrationnel et, donc, d'avoir deux intentions incompatibles. Je *peux* avoir en même temps l'intention de laisser mon auto à ma compagne et l'intention d'aller à la bibliothèque en auto, bien que je sois alors coupable d'irrationalité. Or, comme je vais le montrer dans le prochain paragraphe, si l'intention est identifiée au désir dominant, il est *impossible* d'avoir en même temps des intentions incompatibles. Donc, si l'on identifie l'intention à un désir dominant, la façon dont il faut comprendre la contrainte de compatibilité est bien différente de la façon dont il faut la comprendre lorsqu'elle se rapporte à la notion d'intention décrite plus haut. Alors que, dans un cas, la contrainte de compatibilité rend impossible l'incompatibilité entre intentions, dans l'autre cas, l'incompatibilité est parfaitement possible (bien qu'irrationnelle). Essayons

maintenant de comprendre pourquoi l'intention comprise comme désir dominant rend impossible la présence simultanée d'intentions incompatibles.

Selon l'identification de l'intention à un désir dominant, si j'ai l'intention de faire l'action X, alors l'action X fait l'objet d'un désir dominant. Or, si cette action fait l'objet d'un désir dominant, alors, par définition, elle est davantage désirée que les actions qui sont incompatibles avec elle. Mais si l'action X est plus désirée que les actions qui lui sont incompatibles, alors ces dernières actions ne sont pas dominantes. Mais si les actions qui sont incompatibles avec l'action X ne sont pas l'objet de désirs dominants, alors, conformément à l'identification entre intention et désir dominant, ces actions *ne peuvent pas* être visées par des intentions. Donc, selon la thèse réductionniste, si j'ai l'intention de faire l'action X, les actions incompatibles avec cette action *ne peuvent pas* être visées par des intentions.

Alors, clairement, l'identification de l'intention à un désir dominant ne permet pas d'expliquer les quatre caractéristiques qui ont été attribuées aux intentions. Il faut donc compléter cette première tentative de réduction par une nouvelle condition. La seconde tentative est celle-ci : avoir l'intention de faire X, c'est avoir le désir dominant de faire X *et* avoir la croyance que je vais faire X en raison de ce désir. Ainsi, selon cette nouvelle tentative de réduction, avoir l'intention d'aller à la libraire, c'est avoir le désir dominant d'aller à la libraire ainsi que la croyance que, compte tenu de ce désir, je vais aller à la bibliothèque. Est-ce que cette nouvelle condition permet de satisfaire les ambitions réductionnistes? Bien que cette nouvelle condition améliore les choses, elle n'est pas satisfaisante.

Premièrement, elle ne permet de répondre à notre dernière objection : cette nouvelle tentative de réduction rend toujours impossible, et non pas seulement irrationnelle, la présence d'intentions incompatibles.

Deuxièmement, cette tentative de réduction n'explique pas que, lorsque je suis habité par une intention, les choses cessent d'être sujettes à la délibération. Bratman (1987) donne deux exemples pour nous en convaincre.

Supposons que je suis habité par le désir dominant d'aller à la bibliothèque de l'UQAM et que je croie que, compte tenu de ce désir, je vais terminer à la bibliothèque de l'UQAM[14]. Dans ces circonstances, puis-je dire que je cesse de délibérer à propos de la possibilité d'aller à la bibliothèque de l'UQAM? Pas nécessairement. Il est possible que je suspecte que mon désir dominant d'aller à la bibliothèque de l'UQAM résulte de ma tendance workaholique. Et cette idée peut me pousser à continuer à m'interroger sur la pertinence d'aller à la bibliothèque. Je pourrais donc être enclin à délibérer encore sur l'action d'aller à la bibliothèque, quand bien même que je possède le désir dominant d'aller à la bibliothèque et que je continue à croire que, compte tenu de ce désir, je vais terminer à la bibliothèque.

Voici l'autre exemple de Bratman. Supposons que j'ai une envie passagère de manger un morceau de chocolat pour dessert et que cette envie correspond à un désir dominant de manger ce morceau de chocolat. Supposons aussi que, compte tenir de ce désir, je crois que je vais finir par manger ce morceau de chocolat. Mais, prenant conscience de ce désir, je peux me mettre à réfléchir à mon objectif de suivre une diète. Cette réflexion me pousse alors à sauter ce dessert. Selon la deuxième tentative de réduction, j'ai eu une intention passagère de manger ce morceau de chocolat. « But I am inclined to say I had no such intention, for I was never appropriately settled in favor of such dessert. » (Bratman, 1987, p. 20)

Comme je l'ai montré plus tôt lorsque j'ai mis la première tentative de réduction en défaut, les deux autres caractéristiques de l'intention, à savoir que l'intention suscite un raisonnement procédant des fins aux moyens et qu'elle contrôle l'action, ne peuvent pas être satisfaites si l'intention de l'agent n'a pas préalablement mis un terme à la délibération. Or, je viens de montrer que cette nouvelle tentative de réduction (désir dominant + croyance) ne permet pas d'expliquer que l'intention de l'agent met fin à la délibération. Donc, ces deux autres caractéristiques ne sont pas non plus expliquées par cette nouvelle tentative de réduction.

Donc, les quatre caractéristiques attribuées aux intentions ne peuvent pas être expliquées par un complexe de croyances et de désirs. Il faut donc considérer les intentions comme des états mentaux à part entière, non réductibles à un complexe de croyances et de désirs.

---

[14] Évidemment, l'exemple de Bratman ne réfère pas à la bibliothèque de l'UQAM. En fait, cet exemple réfère à la bibliothèque Tanner.

M'inspirant de ce qui a été dit jusqu'ici sur les intentions et m'inspirant de Bratman *et al.* (1988), je vais maintenant décrire l'architecture cognitive qui génère les intentions.

2.2. L'architecture cognitive à la base du plan d'action

Pour établir cette architecture cognitive, deux des quatre caractéristiques énoncées ci-dessus seront particulièrement importantes. Ces deux caractéristiques sont celles qui concernent le raisonnement de l'agent : l'intention sert d'input à un raisonnement procédant des fins aux moyens et l'intention est un « filtre » de compatibilité. Le fait que ces deux caractéristiques soient particulièrement importantes au sein de l'architecture cognitive n'a d'ailleurs rien d'étonnant. L'architecture cognitive présentée ici définit les processus cognitifs qui sont à la base de nos plans et le raisonnement fait évidemment partie de ces processus cognitifs. Cette architecture est présentée à la figure 2.1.

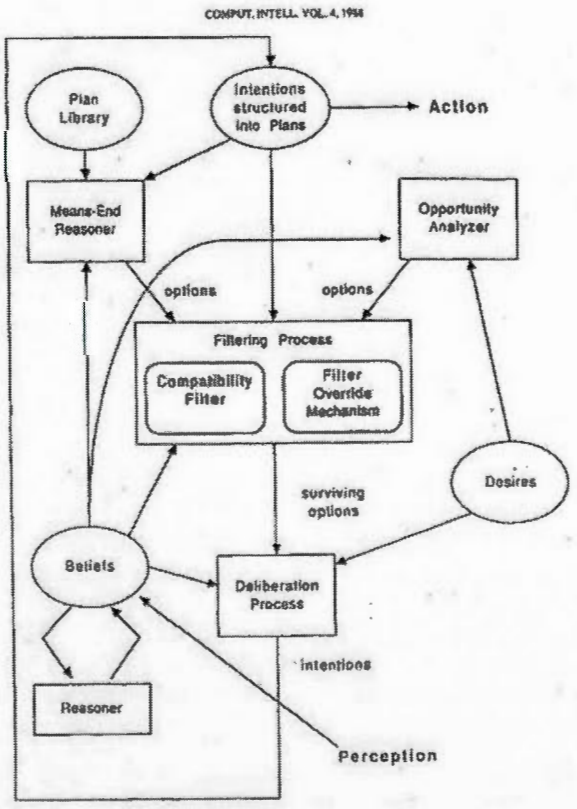


Figure 2.1 L'architecture cognitive. Figure tirée de Bratman *et al.* (1988).

Pour expliquer cette architecture cognitive, je supposerai d'abord que l'agent s'est préalablement doté d'une intention. Cette intention préalable est issue d'un processus de délibération. La délibération est un processus qui évalue les actions en fonction des désirs et des croyances de l'agent. L'action retenue est celle qui possède le plus d'utilité attendue (expected utility). Il s'agit donc de développer une architecture cognitive qui explique comment, à partir de cette intention préalable (l'action retenue par le processus de délibération), le système élabore et développe un plan complet.

Ci-dessus, nous avons convenu que l'intention préalable servait d'input à un raisonnement procédant des fins aux moyens (means-end reasoning). Donc, l'architecture cognitive comprend à tout le moins les deux composantes suivantes. Elle comprend une intention préalable, qui constitue un plan partiel. Dans la figure 2.1, cette intention préalable entre dans la bulle « intentions structured into plans ». Et cette dernière composante agit à titre d'input sur une autre composante, qui est en fait un processus où se produit un raisonnement procédant des fins aux moyens. Bratman *et al.* (1988) nomme cette dernière composante « raisonneur procédant des fins aux moyens (means-end reasoner) ». Le raisonneur procédant des fins aux moyens est activé par deux autres composantes de l'architecture : d'une part, par les croyances de l'agent, et, d'autre part, par l'ensemble comprenant les représentations motrices[15] de l'agent (dans le langage de l'IA : l'ensemble des opérateurs). Dans la figure 2.1, l'ensemble des opérateurs entre dans la bulle « Plan Library ». Lorsqu'il est activé par ces trois composantes, le raisonneur procédant des fins aux moyens produit un raisonnement dont les outputs sont des représentations motrices, c'est-à-dire, plus précisément, des

[15] Par l'expression « représentation motrice », je ne renvoie pas nécessairement aux représentations qui ont lieu dans le système moteur. Par cette expression, je réfère simplement à des représentations qui ont pour contenu un mouvement. À ce titre, le système cognitif responsable des intentions de l'agent, qui n'est pas le système moteur, peut avoir de telles représentations. En effet, les intentions représentent les mouvements de l'agent. Je rappelle ici que les intentions représentent le mouvement de manière conceptuelle. Plus précisément, au sommet de la hiérarchie, l'intention possède un contenu fortement conceptuel, détaché du contexte perceptuel. Et plus on descend dans la hiérarchie, plus les intentions sont particulières, sans toutefois perdre complètement leur valeur conceptuelle. Aussi, les intentions du bas de la hiérarchie sont ancrées dans le contexte perceptuel. Ici, il faut noter que, à première vue, l'architecture cognitive présentée ici ne permet pas de distinguer entre ces différentes sortes d'intentions. L'architecture de Bratman *et al.* (1988) a donc ses limites. Une recherche plus poussée devrait tenter de compléter l'architecture de Bratman de manière à expliquer ces différences entre nos intentions.

représentations motrices ou opérateurs qui sont autant de moyens pour réaliser l'intention préalable (la fin). Celles-ci (les représentations motrices ou opérateurs) sont ensuite envoyées vers la composante délibérative du système, où l'un des moyens est sélectionné[16]. L'option retenue est alors envoyée dans la composante du système où prend place le plan d'action; le plan d'action gagne ainsi en « profondeur ».

Cette description de l'architecture cognitive CDI passe sous silence plusieurs composantes importantes. Avant de décrire ces aspects importants, j'aimerais m'attarder sur le raisonneur procédant des fins aux moyens. Plus précisément, j'aimerais comprendre comment cette composante, une fois qu'elle est alimentée par les trois sources susmentionnées, produit les options (les opérateurs) qui sont envoyées à la composante délibérative. Pour comprendre comment ce raisonneur procède pour faire ce qu'il fait, je vais m'appuyer sur le texte de Fikes et Nilsson (1971), *STRIPS : A New Approach to the Application of Theorem Proving to Problem Solving*. Le choix de ce texte est justifié. Premièrement, il est simple et facile d'accès. Deuxièmement, Bratman *et al.* (1988) y réfèrent lorsqu'ils traitent du raisonneur procédant des fins aux moyens. Enfin, troisièmement, même si cet article est assez vieux, il garde de son actualité. En effet, en parlant de cet article, Suchman (2007) soutient que le « [s]ubsequent work on problem solving and plan synthesis consisted in large part in refinement to this basic means-ends strategy [...]. » (Suchman, 2007, p. 53)

2.2.1. Le raisonneur procédant des fins aux moyens

La tâche du raisonneur est de trouver une séquence d'opérateurs qui fait en sorte que les modèles du monde du raisonneur vont se succéder jusqu'à ce que l'on parvienne à un modèle du monde où le but de l'agent (son intention ou sa fin) est atteint[17]. En effet, un opérateur (ou action) a pour propriété de changer ou transformer un modèle du monde en un autre modèle du monde. La séquence d'opérateurs qui permet de passer du modèle du monde actuel à un

---

[16] L'architecture cognitive de la figure 2.1 interpose un processus de filtration entre le raisonneur procédant des fins aux moyens et le processus de délibération. Pour l'instant, je passe sous silence ce processus de filtration.

[17] Le raisonneur part du modèle du monde actuel. Comme ce modèle du monde est celui auquel *croit* le système, on comprend pourquoi notre architecture connecte la composante « croyance » au raisonneur procédant des fins aux moyens.

modèle du monde où le but est atteint constitue le moyen par lequel l'agent réalise son intention.

Donc, le raisonneur se pose un problème : Quelle séquence d'opérateurs permet de passer du modèle du monde actuel à un modèle du monde où le but est atteint? Le problème du raisonneur est donc défini par trois éléments : 1) l'intention ou but de l'agent, 2) le modèle du monde actuel, et 3) la « banque » d'opérateurs qui sont à la disposition du raisonneur. Autrement dit, le raisonneur doit être alimenté par ces trois types d'input. C'est précisément ce qui est indiqué à la figure 2.1. L'intention est représentée dans la bulle « Intentions structured into Plans », le modèle du monde actuel, dans la bulle « Beliefs », et la « banque » d'opérateurs, dans la bulle « Plan Library ». Dans la prochaine section, je vais montrer comment l'architecture cognitive représente ces trois types d'input.

2.2.1.1 La représentation des trois inputs du raisonneur

Le but est représenté par une expression bien formée (ebf)[18]. Ainsi, dans un exemple de Fikes et Nilsson (1971), le but de l'agent consiste à placer les boîtes A et B à l'endroit a, et celui-ci est représenté par la ebf suivante :

$$AT(A, a) \wedge AT(B, a)^{19}$$

Deuxièmement, le problème du raisonneur est défini par un modèle du monde actuel. En IA, le modèle actuel du monde est aussi représenté par une ebf. Ainsi, si le monde se caractérise par ceci que le robot (l'agent) est à l'endroit a et que les boîtes B et C sont aux endroits b et c respectivement, le modèle du monde est représenté par ces ebfs :

$$ATR(a)^{20}$$

$$AT(B, b)$$

$$AT(C, c)$$

---

[18] L'expression « expression bien formée » est la traduction de l'expression anglaise « well formed formula ou wff ».

[19] Le symbole « AT » signifie la relation « à ». Il renvoie au mot anglais « at ».

[20] Le symbole « ATR » est un prédicat qui attribue à un endroit la propriété d'être occupé par un robot.

Le raisonneur doit aussi tenir compte des opérateurs. Les opérateurs sont définis à l'aide de deux composantes. L'opérateur est d'abord défini par les effets qu'il produit. « The effects of an operator are simply defined by a list of wffs that must be added to the model and a list of wffs that are no longer true and therefore must be deleted. » (Fikes et Nilsson, 1971, p. 192) La deuxième composante de l'opérateur spécifie les conditions sous lesquelles l'opérateur peut être appliqué. Ces conditions d'applicabilité, ou préconditions, sont également formulables comme des ebfs.

Avant de décrire plus précisément les autres éléments de cette architecture, je vais tenter de comprendre comment le raisonneur parvient à résoudre son problème. Autrement dit, je vais montrer comment le raisonneur parvient à trouver une séquence d'opérateurs qui lui permet de passer d'un modèle du monde actuel à un modèle du monde où le but est atteint.

2.2.1.2. La recherche d'une séquence d'opérateurs à l'aide du GPS

Le raisonneur résout son problème en adoptant la stratégie mise de l'avant par le GPS (general problem solver). Le GPS a été proposé pour la première fois par Newell *et al.* (1957). Cependant, Fikes et Nilsson (1971) appuient leur description du GPS sur l'article de Ernst et Newell (1969). Il est aussi important de noter que, malgré son ancienneté, le GPS est toujours d'actualité. Par exemple, Simon (1996) continue de défendre sa pertinence.

Pour ce qui est de la résolution d'un problème très simple, la stratégie de recherche peut procéder de la manière suivante. On applique tous les opérateurs au modèle actuel du monde de manière à créer un ensemble de modèles successeurs. Et on continue d'appliquer tous les opérateurs à chacun de ces modèles successeur ainsi qu'aux autres modèles qui vont résulter de ces applications d'opérateurs. De cette manière, on peut espérer arriver à un modèle du monde où le but est atteint, c'est-à-dire, plus précisément, à un modèle du monde duquel on peut démontrer, à l'aide d'un « theorem-prover », la ebf correspondant au but. Cette stratégie de recherche peut donner des fruits lorsque l'agent a peu d'opérateurs à sa disposition. Mais l'agent qui nous intéresse contient beaucoup d'opérateurs, si bien que, telle que définie ici, la recherche d'une séquence d'opérateurs peut devenir impraticable.

La stratégie de recherche appelée « GPS », qui est celle que Fikes et Nilsson adoptent, est, quant à elle, praticable.

Le GPS extrait d'abord des « différences » entre le modèle actuel du monde et le but ou intention de l'agent. Les « différences » entre le modèle actuel du monde et le but sont déterminées de la manière suivante. À l'aide d'un démonstrateur de théorème (theorem-prover), on tente de démontrer le but à partir du modèle actuel du monde, et les différences entre ce modèle et le but correspondent aux éléments incomplets de la preuve. Plus la preuve est incomplète, plus les différences sont nombreuses.

Ensuite, une fois que ces différences ont été établies, le raisonneur tente de trouver des opérateurs qui réduiront ces différences. Les opérateurs qui réduisent ces différences sont les opérateurs dont les effets permettent de compléter la preuve incomplète. Une fois que les opérateurs pertinents ont été trouvés, il s'agit de les appliquer.

Pour appliquer un opérateur au modèle actuel du monde, il faut considérer la condition d'applicabilité ou précondition de l'opérateur. Si, grâce au démonstrateur de théorème, la précondition de l'opérateur peut être démontrée à partir du modèle actuel du monde, l'opérateur est appliqué. À la suite de l'application, on considère (évidemment off-line) le modèle du monde résultant. On tente alors de déterminer les différences entre ce nouveau modèle du monde et le but, et, s'il y a des différences, on tente d'appliquer de nouveaux opérateurs, jusqu'à ce qu'on arrive à un modèle du monde à partir duquel le but peut être prouvé ou démontré.

Mais qu'arrive-t-il si un opérateur pertinent a été trouvé (c'est-à-dire un opérateur dont les effets permettent de réduire les différences ou, autrement dit, permettent de compléter la preuve incomplète) et que cet opérateur ne s'applique pas au modèle du monde? Autrement dit, qu'arrive-t-il lorsque les préconditions de l'opérateur pertinent ne peuvent pas être démontrées à partir du modèle du monde? Les choses se compliquent alors légèrement.

Les préconditions de l'opérateur pertinent constituent alors le sous-but du raisonneur. Le raisonneur établit les différences entre le modèle du monde et les préconditions de l'opérateur et il tente de trouver des opérateurs qui vont réduire ces différences de manière à produire un nouveau modèle du monde à partir duquel l'opérateur qui était pertinent en premier lieu pourra être appliqué. Et cet autre opérateur, celui qui permet d'éliminer ou, à tout le moins, de réduire la différence entre le modèle du monde et les préconditions de l'opérateur qui était pertinent en premier lieu, peut lui-même ne pas s'appliquer à ce modèle du monde. Les

préconditions de cet opérateur serviront alors de (sous-)sous-but au raisonneur. Lorsque les sous-buts s'emboîtent outre mesure les uns dans les autres, le raisonneur contient des mécanismes heuristiques qui font en sorte que les séquences d'opérateurs correspondant à ces sous-buts ne sont pas poursuivies. Les mécanismes heuristiques répondent à des critères qu'il est inutile d'exposer ici.

Avant de passer à la prochaine étape, qui est de définir les autres composantes de l'architecture cognitive de la figure 1, j'aimerais montrer à l'aide d'un exemple comment, à travers son processus de recherche, le GPS trouve une séquence d'opérateurs permettant de satisfaire un but. Cet exemple est tiré de l'article de Fikes et Nilsson susmentionné.

La recherche de la séquence d'opérateur peut être représentée à l'aide d'un arbre de recherche (search tree). L'arbre de recherche de Fikes et Nilsson est donné dans la figure 2.2.

Mon but maintenant est de donner une explication de cet arbre de recherche. Comme on le constate, les noeuds de cet arbre de recherche possèdent deux composantes majeures : un modèle du monde, représenté par la lettre M, et des buts, représentés par la lettre G. Dans l'arbre de recherche de la figure 2.2, le GPS part avec un modèle du monde M_o et un but G_o. Le noeud représentant cet état initial du GPS est le noeud qui est situé au sommet de l'arbre. Le but G_o ne peut pas être démontré à partir de M_o. Le GPS doit donc trouver des opérateurs qui vont réduire la différence qui existe entre M_o et G_o. Le GPS trouve deux opérateurs. Le premier opérateur a pour précondition le sous-but G_a et le second a pour précondition le sous-but G_b. Ces sous-buts sont placés en premier dans la liste des buts. Donc, on voit ici apparaître deux nouveaux noeuds. L'un est le noeud de l'embranchement de gauche, (M_o, (G_a, G_o)), où la précondition de l'opérateur A, le sous-but G_a, est en premier dans la liste des buts. L'autre node correspond à l'embranchement de droite, où la précondition de l'opérateur B, le sous-but G_b, est en premier dans la liste des buts. Examinons d'abord l'embranchement de gauche.

On constate que le sous-but G_a est démontrable à partir du modèle du monde M_o. Le GPS peut donc appliquer l'opérateur correspondant à ce sous-but (précondition), c'est-à-dire l'opérateur A. On constate dans l'arbre de recherche que l'opérateur est appliqué et qu'il en résulte un nouveau modèle du monde, M_1. Il s'agit alors de voir si le but initial, G_0, est démontrable à partir de ce nouveau modèle du monde. Il semble que non, puisqu'un nouveau

noeud, doté d'un but additionnel, fait suite au noeud (M_1, G_o). On constate alors que le GPS ne continue pas la recherche. Sur la base de certaines heuristiques, le GPS a dû considérer que cette recherche ne valait pas la peine. Considérons maintenant le second embranchement, l'embranchement de droite, où la recherche du GPS est menée à terme.

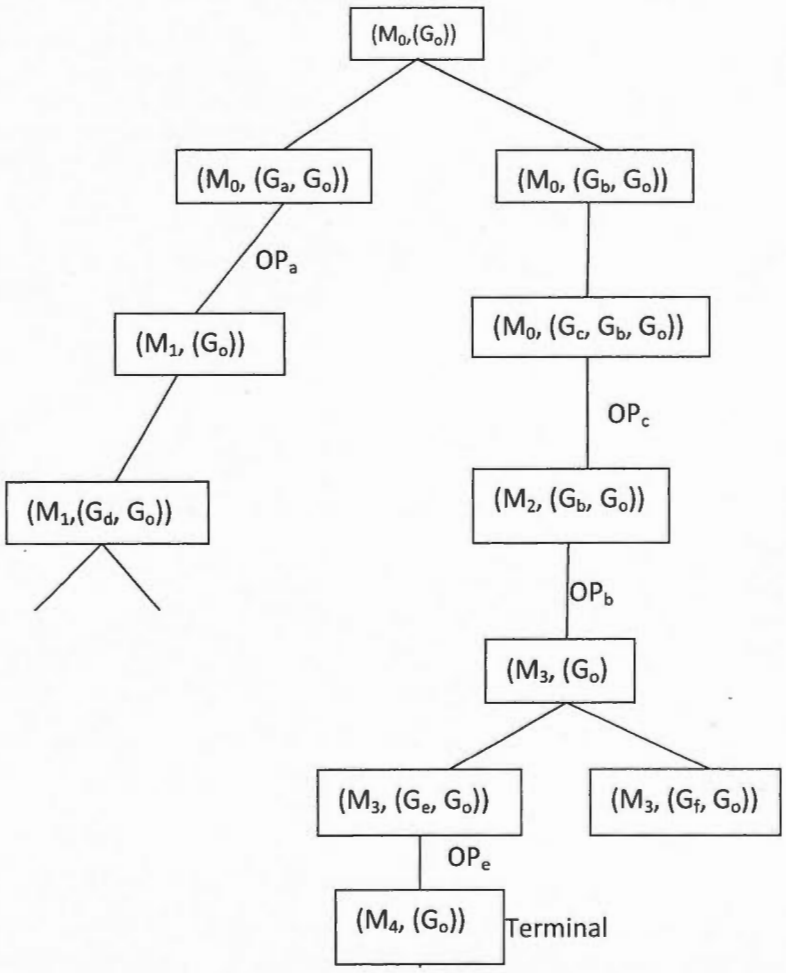


Figure 2.2 Un arbre de recherche. Tiré de Fikes et Nilsson (1971).

Dans cet embranchement, l'opérateur B est considéré afin de réduire la différence qui existe entre M_0 et le but G_o. Un nouveau noeud est créé où le sous-but G_b est mis en premier dans la liste. Il s'agit alors de voir si G_b est démontrable à partir de M_0. Il semble que non, puisqu'un autre (sous-)sous-but est ajouté dans le noeud suivant, (M_0, (G_c, G_b, G_o)). Cela signifie que le GPS doit trouver un opérateur qui réduit la différence existant entre M_0 et G_b et

que la précondition correspondant à ce nouvel opérateur est G_c. Le GPS doit ensuite voir si G_c est démontrable à partir de M_0. Il semble que oui puisque l'opérateur C est appliqué. Un nouveau noeud est donc créé dans lequel on retrouve un nouveau modèle (M_2) du monde et où le (sous)sous-but G_c est éliminé. Il s'agit ensuite de voir si le premier sous-but dans la liste de buts est démontrable à partir du nouveau modèle du monde, M_2. Il semble que oui, puisque l'opérateur correspondant à ce sous-but, l'opérateur B, est appliqué. Compte tenu de l'application de cet opérateur, un nouveau modèle du monde est créé et le sous-but G_b est éliminé. Le noeud (M_3, G_o) est donc créé. Il s'agit ensuite de voir si G_o est démontrable à partir de M_3. Il semble que non puisqu'un nouveau noeud est ajouté où l'on retrouve un nouveau sous-but (G_e), qui correspond à la précondition d'un opérateur qui réduit la différence entre M_3 et G_o. Le GPS tente alors de voir si ce nouveau sous-but est démontrable à partir de M_3. Il semble que oui, puisque l'opérateur E est appliqué. Un nouveau noeud est créé dans lequel on retrouve un nouveau modèle du monde (M_4) et, comme ce nouveau noeud est terminal, cela signifie que G_o, le but initial, est démontrable à partir de M_4. On peut donc atteindre son but à partir de la séquence d'opérateurs (OP_c, OP_b, OP_e). C'est cette séquence d'opérateurs qui sera envoyée vers la composante délibérative.

Il est maintenant temps de décrire le reste de l'architecture cognitive.

2.2.2 Les autres composantes de l'architecture cognitive

Le raisonneur procédant des fins aux moyens trouve des séquences d'opérateurs qui permettent de réaliser une fin (l'intention préalable). Ces séquences d'opérateurs ou ces différents moyens pour réaliser l'intention préalable sont considérés par le système comme autant d'options qui, si elles passent à travers le processus de filtration (défini plus bas), sont envoyées vers le système délibératif. Le système délibératif, sur la base des croyances et des désirs de l'agent, décide alors de l'option à privilégier et l'option sélectionnée acquiert alors le statut d'intention; elle est envoyée dans la composante « Intentions structured into plans » de la figure 2.1.

Mais, comme le montre l'architecture cognitive de la figure 1, les options qui sont envoyées vers le système délibératif ne sont pas toutes issues d'un raisonnement procédant

des fins aux moyens; des options peuvent provenir de l'analyseur d'opportunités (opportunity analyser).

2.2.2.1. L'analyseur d'opportunités

Un événement inattendu peut se produire et cet événement peut se présenter comme une opportunité pour satisfaire un désir ou comme une opportunité pour éviter une menace. Lorsque c'est le cas, l'analyseur d'opportunités émet une option (une action) qui permet de satisfaire le désir ou d'éviter la menace.

Voici un exemple où l'événement se présente comme une opportunité permettant de satisfaire le désir de l'agent. Un joueur de tennis planifie (a l'intention préalable) de jouer en fond de terrain et son adversaire se retrouve dans une situation très défavorable – par exemple, il est très loin derrière la ligne de fond. Cette situation se présente comme une opportunité pour satisfaire le désir du joueur d'en finir avec le point. Face à cette opportunité, l'analyseur d'opportunités émet l'option « Exécute un amorti ».

Un exemple d'opportunité qui permet d'éviter une menace serait le suivant. Je suis en train de jouer au badminton dans un gymnase et l'alarme de feu sonne. Cet événement se présente comme une opportunité pour éviter la menace du feu. Face à cette opportunité, l'analyseur d'opportunités émet l'option « Quitte ce terrain ».

L'option du joueur de tennis (jouer un amorti) et celle du joueur de badminton (quitter cet immeuble) ne sont apparemment pas des moyens permettant de réaliser une intention (une fin). En effet, l'option de jouer un amorti n'est pas un moyen permettant de jouer en fond de terrain et l'option de quitter l'immeuble n'est pas un moyen permettant de jouer au badminton. En fait, dans le cas du joueur de tennis, si le système délibératif décide de sélectionner cette option, cette option vient supplanter l'intention de jouer en fond de terrain. C'est aussi le cas pour le joueur de badminton, puisque, si elle est adoptée, l'option de quitter l'immeuble supplante l'intention de jouer au badminton. Bratman *et al.* (1988) affirment que l'option de l'analyseur d'opportunités est le moyen d'*aucune* fin (intention) déjà visée. En effet, en parlant de l'analyseur d'opportunités, ils affirment que :

> Changes in the agent's environment may lead to changes in her beliefs, which in turn may result in her considering new options that are not means to any already intended end. (Bratman *et al.*, 1988, p. 350)

Dans l'exemple du joueur de badminton, l'option est effectivement le moyen d'aucune fin, mais, dans nombre de cas, l'option peut être le moyen d'une fin, même si, au lieu d'être produite par le raisonneur procédant des fins aux moyens, elle (l'option) est le produit de l'analyseur d'opportunités. En effet, dans de nombreux cas, l'option émise par l'analyseur d'opportunités s'oppose à une intention qui est le moyen d'une fin supérieure (c'est-à-dire, située plus haut dans la hiérarchie composant le plan) et l'option peut être un moyen relativement à cette fin supérieure. Reprenons l'exemple du joueur de tennis. L'intention de jouer en fond de terrain est un moyen par rapport à l'intention de gagner le match, et, relativement à cette fin supérieure, l'option de jouer un amorti (l'option émise par l'analyseur d'opportunités) est aussi un moyen. Donc, pour caractériser adéquatement le rapport de l'option au plan, il ne faut pas dire, comme le font Bratman *et al.* (1988), que l'option de l'analyseur d'opportunités est le moyen d'aucune fin, mais il faut simplement dire que cette option peut supplanter l'une des intentions composant le plan et que, lorsque l'intention supplantée est un moyen relativement à une fin supérieure, l'option devient elle aussi un moyen relativement à cette fin supérieure. Par ailleurs, si Bratman et ses collègues étaient cohérents avec l'exemple qu'ils présentent dans leur texte, ces derniers devraient aussi reconnaître que l'option émise par l'analyseur d'opportunités peut être un moyen relativement à une fin supérieure. Dans leur exemple, Bratman *et al.* (1988) considère un robot, Rosie, qui a pour fin de réparer un appareil vidéo défectueux. Après avoir raisonné sur les moyens, Rosie décide de remplacer une pièce de l'appareil. Lorsqu'elle se dirige vers l'appareil pour effectuer le remplacement, Rosie constate que le contraste de l'appareil est fermé. Cette observation se présente comme une opportunité pour ouvrir le contraste. Cette option supplante l'intention de remplacer la pièce de la vidéo. Dans cet exemple, l'option d'ouvrir le contraste s'oppose effectivement à l'intention de remplacer la pièce, mais elle reste un moyen relativement à la fin supérieure qui est de réparer l'appareil vidéo.

L'option émise par l'analyseur d'opportunités peut-elle être compatible avec la totalité des intentions préexistantes composant le plan. Cela est évidemment possible. Par exemple, je peux avoir l'intention d'aller chercher mes neveux en après-midi et, en cours de route,

l'opportunité de m'acheter un livre à bon marché peut se présenter à moi, si bien que l'analyseur d'opportunités émet l'option « Achète ce livre rare ». Cette option est compatible avec mon intention de chercher mes neveux et, possiblement, avec toutes les autres intentions composant mon plan. Donc, il faut retenir que l'option émise par l'analyseur d'opportunités ne s'oppose pas toujours aux intentions préalables du plan.

L'option émise par l'analyseur d'opportunités est fonction de deux éléments : les croyances et les désirs de l'agent. En effet, pour que l'analyseur d'opportunités émette son option, il doit avoir certaines croyances à propos du monde, Par exemple, pour émettre l'option « Exécute un amorti », l'analyseur d'opportunités doit *croire* que le joueur adverse est très loin derrière la ligne de fond. De même, pour émettre l'option de quitter l'immeuble, le joueur de badminton doit croire qu'une alarme de feu sonne. Et l'option, de toute évidence, est aussi fonction des désirs de l'agent; si le joueur de tennis n'avait pas le désir d'en finir au plus tôt avec le point, l'analyseur d'opportunités n'émettrait pas l'option de jouer un amorti. De même, si le joueur de badminton n'avait pas le désir de sauver sa peau, l'analyseur d'opportunités n'émettrait pas l'option de quitter l'immeuble. Donc, tel que posé ci-dessus, l'analyseur d'opportunités doit être connecté aux croyances et aux désirs de l'agent. Ceci est conforme à ce qui est indiqué dans la figure 2.1.

Il reste maintenant à décrire le mécanisme qui permet à l'analyseur d'opportunités d'émettre des options. Plus précisément, il s'agit de comprendre comment, sur la base de croyances et de désirs, l'analyseur d'opportunités émet des options vers le système délibératif. Bratman *et al.* (1988) ne décrivent pas le mécanisme par lequel l'analyseur d'opportunités émet des options. J'ai trouvé chez Simon (1996) un mécanisme qui pourrait convenir à l'analyseur d'opportunités. Je vais maintenant décrire ce mécanisme.

Je comprends l'analyseur d'opportunités comme un *système de production* (Simon, 1996). Un système de production est un système de règles ou productions qui sont indépendantes les unes des autres, c'est-à-dire un système non hiérarchique de règles. Les règles, ou productions, ont la forme « Si..., alors... ». L'antécédent de la règle est nommé « condition » et le conséquent « action ». Selon Simon (1996), la condition d'une règle peut être de plusieurs types. Pour adapter les productions (ou règles) à notre architecture cognitive, je

considère que l'antécédent d'une règle correspond à une croyance perceptuelle. Le conséquent d'une règle, ou action, est un opérateur.

J'ai dit ci-dessus que les règles étaient indépendantes les unes des autres. Cela signifie que, pour activer une règle, il suffit que la condition ou antécédent de la règle soit satisfait; autrement dit, il suffit que l'agent ait une croyance qui corresponde à l'antécédent. Lorsque, à la suite de la satisfaction de l'antécédent, une règle est activée, le conséquent est aussi activé. L'activation du conséquent implique que l'analyseur d'opportunités émet comme option l'action correspondant au conséquent.

Afin d'adapter l'analyseur d'opportunités (maintenant défini comme un système de productions) à notre architecture cognitive, nous allons considérer que les productions ne sont pas seulement activées par des croyances perceptuelles; les productions sont aussi *pré*activées par les désirs de l'agent. Plus précisément, le désir de l'agent « prépotentialise » (prepotentiates) un ensemble de règles et ce sont seulement ces règles « prépotentialisées » qui peuvent être activées lorsque leur antécédent est satisfait par une croyance perceptuelle.

Ainsi, le désir de gagner le point au tennis « prépotentialise » un certain nombre de règles, au sein desquelles il pourrait y avoir la règle « Si l'adversaire est loin derrière la ligne de fond, alors exécute un amorti ». Parce que cette règle a été prépotentialisée, elle peut être activée. Si l'agent croit que son adversaire est loin derrière la ligne de fond, la règle susmentionnée s'active : l'option de jouer un amorti est envoyée au système délibératif.

À ce stade, on pourrait se poser cette question : Quel ensemble de règles le désir prépotentialise-t-il? Le désir prépotentialise évidemment les règles qui ont trait au désir. Ainsi, le désir de gagner le point au tennis prépotentialise les règles qui sont relatives au tennis.

On pourrait également s'interroger sur l'intérêt d'avoir des désirs qui prépotentialisent un ensemble de règles au détriment du reste des règles. En effet, ne pourrait-on pas envisager un système de production où *toutes* les règles sont prépotentialisées[21]? Cela pourrait occasionner des problèmes, car des règles différentes ayant le même antécédent pourraient être activées en même temps. Bien sûr, il est possible que certaines des règles prépotentialisées par le désir

[21] Cela revient à dire : si aucune règle n'a besoin d'être prépotentialisée pour pouvoir être activée.

aient un antécédent identique. Mais le fait que le désir prépotentialise un ensemble de règles et laisse de côté le reste des règles diminue grandement le nombre de règles pouvant avoir le même antécédent. Donc, comme l'action d'une règle est activée lorsque son antécédent est activé et comme c'est au système délibératif que revient la tâche de choisir une action (ou option) au sein des actions émises par l'analyseur d'opportunités, le désir diminue le travail délibératif du système délibératif.

Jusqu'ici, nous avons supposé que les options issues de l'analyseur d'opportunités ainsi que celles provenant du raisonneur procédant des fins aux moyens étaient directement envoyées au système délibératif. Comme le montre l'architecture de la figure 2.1, cela n'est pas tout à fait juste. Les options issues de l'analyseur d'opportunités passent d'abord à travers un processus de filtration.

2.2.2.2. Le processus de filtration

Dans l'architecture cognitive de la figure 1, le processus de filtration contient deux composantes : le filtre de compatibilité et le mécanisme de dérogation au filtre. C'est dans la première composante que l'information est d'abord acheminée. Plus précisément, les options issues du raisonneur procédant des fins aux moyens ainsi que celles qui proviennent de l'analyseur d'opportunités se dirigent d'abord vers le filtre de compatibilité. Mais le filtre de compatibilité ne reçoit pas seulement de l'information provenant du raisonneur et de l'analyseur d'opportunités. La fonction du filtre de compatibilité est de faire en sorte que les options soient compatibles avec les plans et les croyances de l'agent. Donc, en plus de recevoir les options provenant du raisonneur et de l'analyseur d'opportunités, le filtre de compatibilité doit recevoir de l'information à propos des plans de l'agent ainsi que de l'information à propos des croyances de l'agent. Ceci est conforme à l'architecture de la figure 2.1.

Quand je dis que la fonction du filtre de compatibilité est de faire en sorte que l'option soit compatible avec les plans de l'agent, je dis simplement ceci : Si le raisonneur ou l'analyseur émet une option qui est incompatible avec les plans préalablement établis, cette option ne sera pas envoyée au système délibératif. Et en disant qu'il faut que le plan et l'option soient compatibles avec les croyances de l'agent, je dis, par exemple, que si l'agent a planifié de

laisser son auto à sa compagne et que le raisonneur émet l'option d'aller en auto à la bibliothèque de l'UQAM, cette option ne sera pas envoyée au système délibératif si l'agent *croit* qu'il possède une seule auto.

Les options qui passent avec succès à travers le filtre de compatibilité sont directement envoyées au système délibératif. Et j'ai dit à l'instant que les options qui ne passaient pas à travers le filtre de compatibilité n'étaient pas envoyées au système délibératif. Cela n'est pas tout à fait juste. L'option qui ne passe pas avec succès à travers le filtre de compatibilité est en fait envoyée vers la deuxième composante du processus de filtration, le mécanisme de dérogation au filtre, et si l'option passe avec succès à travers ce mécanisme, elle peut ensuite être considéré par le système délibératif. En fait, si l'option passe à travers le mécanisme de dérogation au filtre, l'intention qui est incompatible avec l'option est reconsidérée; le système délibératif doit soupeser l'option ainsi que l'intention qui est incompatible avec cette option et, à l'issue de ce processus délibératif, décider lequel des deux sera retenu et envoyé vers la composante de l'architecture où se trouvent les plans de l'agent.

Le fait qu'une option passe ou non à travers le mécanisme de dérogation au filtre dépend de la sensibilité du mécanisme aux problèmes et aux opportunités. La sensibilité du mécanisme de dérogation aux problèmes et aux opportunités est fonction des « [...] conditions under which some portions of her existing plans is to be suspended and weighed against some other option. » (Bratman *et al.*, 1988, p. 352) Plus ces conditions sont importantes, plus le mécanisme de dérogation est sensible (il laisse passer plus d'options vers le système délibératif); plus ces conditions sont négligeables, moins le mécanisme est sensible (il laisse passer moins d'options vers le système délibératif). Si la sensibilité du mécanisme de dérogation est telle que le mécanisme répond souvent aux problèmes et aux opportunités (il laisse passer beaucoup d'options vers le système délibératif), Bratman *et al.* (1988) considère que l'agent est prudent. Si le mécanisme de dérogation répond à peu de problèmes et d'opportunités, ils qualifient l'agent d'audacieux. Un agent peut être trop prudent ou trop audacieux. Il est trop prudent s'il répond au moindre problème ou opportunité, et il est trop audacieux s'il ne répond à presque aucun problème ou opportunité.

J'ai maintenant terminé l'explication de l'architecture cognitive. Il est maintenant temps de mettre cette architecture en relation avec la question de l'expertise.

2.3. Comment expliquer le passage vers l'expertise

Au chapitre 1, j'ai posé que l'action experte était une action où l'agent s'appuyait sur des plans très généraux, sans « profondeur ». Or, l'architecture cognitive qui vient d'être décrite suggère l'existence d'un agent qui, à travers le raisonneur procédant des fins aux moyens, élabore des plans très complets (dotés de beaucoup de « profondeur »). En effet, nous avons vu que la production d'intentions prenait terme uniquement lorsque le raisonneur procédant des fins aux moyens atteignait le niveau des habiletés et des habitudes. La question à se poser est donc celle-ci : cette architecture peut-elle supporter un agent qui se satisfait d'intentions très générales et qui, donc, se présente comme un agent réactif? Il faut répondre par l'affirmative à cette question. Mais avant de montrer que cette architecture cognitive supporte des agents réactifs, je vais d'abord montrer que, lorsqu'elle supporte un agent planificateur[22], cette architecture n'est pas en mesure de rendre compte de l'expertise.

2.3.1. Pourquoi l'agent planificateur n'est pas un agent expert?

L'expert est celui qui est en mesure de gérer les événements inattendus. Gérer les événements inattendus dans un environnement *statique* ne pose pas vraiment problème. L'agent a alors tout son temps pour résoudre le problème qui lui est posé par l'événement inattendu. En fait, c'est dans un environnement fortement dynamique que la capacité de l'agent à gérer des événements inattendus est mise à l'épreuve. Dans un environnement fortement dynamique, le monde change très rapidement, si bien que le temps pris par l'agent pour gérer l'événement inattendu – c'est-à-dire le temps pris pour arriver à une décision – peut s'étendre au point que la « solution » (la décision ou intention) arrive trop tard. Donc, dans un environnement fortement dynamique, l'expertise se mesure plus particulièrement par la capacité de l'agent à gérer des événements inattendus au sein d'un environnement fortement dynamique. Ci-dessus, nous avons circonscrit le domaine des événements inattendus : ce sont les événements que gère l'analyseur d'opportunités. Donc, un événement inattendu est ou bien une opportunité qui permet de satisfaire un désir ou bien une opportunité qui se présente comme une menace. Je vais maintenant montrer que si l'agent est

---

[22] Un agent est planificateur lorsqu'il fait reposer son action sur des plans dotés de beaucoup de « profondeur ».

planificateur, l'architecture présentée ci-dessus ne permet pas de gérer les opportunités au sein d'un environnement fortement dynamique.

Considérons une opportunité qui se présente comme une menace. Je suis un joueur de tennis *droitier* qui a l'intention de jouer un revers et la balle est retournée considérablement à ma droite. Cet événement, c'est-à-dire le fait que la balle est retournée considérablement à ma droite, se présente à moi comme une menace. Que fait l'architecture cognitive présentée ci-dessus lorsqu'un tel événement se produit? Nous avons vu que l'analyseur d'opportunités émet une option. Disons qu'il émet l'option de jouer un coup droit. Cette option est alors envoyée vers le processus de filtration. Elle passe d'abord à travers le filtre de compatibilité. Comme cette option est incompatible avec l'intention de jouer un revers, elle n'est pas immédiatement envoyée au processus de délibération; plutôt, elle est envoyée vers le mécanisme de dérogation au filtre. Supposons que l'option de jouer un coup droit passe avec succès à travers le mécanisme de dérogation au filtre. Elle est alors envoyée au processus de délibération, où elle sera comparée avec l'option de jouer un revers. Ce n'est qu'à l'issue de cette délibération que l'option sera acceptée et envoyée vers le plan de l'agent. Donc, comme on le voit, l'adoption de cette nouvelle intention (l'intention de jouer un coup droit) prend déjà un temps considérable. En fait, il se pourrait que le temps qui est nécessaire pour adopter cette nouvelle intention soit tel que cette intention se trouve « dépassée » par les événements. Et les choses pourraient être encore pires. Il n'est pas impossible que l'option émise par l'analyseur d'opportunités soit une option générale, c'est-à-dire une option qui ne soit pas assez spécifique pour conduire directement à une action. Cela signifierait que cette option, après être passée par le processus dont il vient d'être question et, donc, après avoir acquis le statut d'intention, suscite un raisonnement procédant des fins aux moyens, cela en vue de trouver une intention qui sera un moyen pour cette intention générale. Le raisonneur tentera donc de trouver de nouvelles options. Si ces options ne sont pas incompatibles avec les autres plans de l'agent, elles sont ensuite envoyées au système délibératif. Le système délibératif, sur la base de ses croyances et désirs, soupèsera alors ces différents moyens. Il en choisira une et il l'enverra composer le plan. Si cette nouvelle intention est assez spécifique, elle se traduira par une action; si elle est trop générale, elle initiera un nouveau raisonnement procédant des fins aux moyens, et le même cycle recommencera. Afin de nous rendre la tâche plus difficile, supposons que, dès le départ, le moyen trouvé par le raisonneur soit

suffisamment spécifique pour immédiatement entraîner une action motrice. Même dans ces circonstances, il est clair que, dans un environnement fortement dynamique, les étapes qui ont mené à cette nouvelle intention (moyen) prennent trop de temps : durant l'exécution de ces étapes, le monde a le temps de changer de manière à rendre inappropriée l'intention (le moyen) à laquelle est parvenue l'architecture cognitive. En fait, les étapes qui prennent du temps sont les étapes du raisonnement procédant des fins aux moyens ainsi que l'étape délibérative.

2.3.2. La « réactivité » de notre architecture cognitive

Le point important est celui-ci : lorsqu'une intention contrôle des perceptions d'affordances, cette intention cesse de servir d'input pour un raisonnement procédant des fins aux moyens. Ainsi, l'intention de marcher contrôle des perceptions d'affordances (les affordances qui prescrivent l'action de marcher) et, pour cette raison, il n'est pas question de trouver les sous-intentions (les moyens) d'avancer la jambe droite et la jambe gauche alternativement. Maintenant, nous avons également dit que, à travers l'apprentissage, les intentions de bas niveau étaient écartées et que, au lieu d'être contrôlées par ces intentions, les perceptions d'affordances étaient prises en charge par des intentions de plus haut niveau[23]. Par exemple, au moment de disputer un point au tennis, le joueur expert est habité par une seule intention, celle de gagner le point, et c'est cette intention qui contrôle toutes les perceptions d'affordances. Donc, puisqu'elle contrôle les perceptions d'affordances, cette intention n'initie pas un raisonnement procédant des fins aux moyens. L'agent se contente d'un plan très général. Autrement dit, notre architecture rend compte de l'agent réactif. L'agent ne perd plus son temps à chercher des intentions qui servent de moyens pour réaliser sa fin ou intention supérieure. Or, s'il est en mesure d'exécuter une action sans perdre de temps à trouver un moyen pour réaliser son intention, l'agent va réagir à temps à l'événement inattendu : l'agent est expert.

---

[23] Au moment d'aborder la question de l'apprentissage, je montrerai comment les intentions de bas niveau sont écartées et comment les perceptions d'affordances sont prises en charge par des intentions plus générales.

Ce qui vient d'être dit présuppose que des intentions très générales peuvent contrôler les perceptions d'affordances. Comme je l'ai dit au chapitre 1, je rendrai compte de cette possibilité plus tard dans ma thèse (au chapitre 9).

2.4. Résumé

Dans la première partie de ce chapitre, j'ai montré, à partir de Bratman (1987), que les intentions étaient des états mentaux à part entière, c'est-à-dire des états mentaux qui n'étaient pas réductibles à des croyances ou à des désirs. Pour montrer cela, j'ai d'abord porté mon attention sur les intentions et j'y ai dégagé quatre caractéristiques. Ces quatre caractéristiques sont : 1) les intentions contrôlent nos actions; 2) nos intentions mettent fin à la délibération; 3) elles servent d'input à un raisonnement procédant des fins aux moyens; et 4) elles imposent un « filtre » de compatibilité. Ensuite, pour montrer que les intentions n'étaient pas réductibles à des croyances ou à des désirs, j'ai présenté deux tentatives de réduction et j'ai montré que ces deux tentatives ne parvenaient pas à expliquer ces quatre caractéristiques.

La première tentative de réduction identifiait l'intention de faire X à un désir dominant de faire X. Pour les raisons suivantes, cette tentative de réduction n'explique pas les quatre caractéristiques des intentions. Premièrement, un désir dominant ne met pas fin à la délibération (2ème caractéristique). Il suffit de penser à l'alcoolique pour s'en persuader. Celui-ci a le désir dominant de boire et, malgré ce désir, son esprit peut être le théâtre d'un combat furieux entre son désir de rester sobre et son désir de s'enivrer. Deuxièmement, le désir dominant ne sert pas d'input à un raisonnement procédant des fins aux moyens (3ème caractéristique) : le désir dominant ne met pas fin à la délibération et, pour essayer de trouver des moyens pour réaliser son intention (en l'occurrence, son désir dominant), il faut avoir mis un terme à la délibération. Troisièmement, le désir dominant ne contrôle pas l'action (1ère caractéristique), car il ne met pas fin à la délibération et, pour contrôler l'action, il faut avoir mis un terme à la délibération. Enfin, quatrièmement, le désir dominant n'impose pas une relation de compatibilité qui soit comparable à la relation de compatibilité imposée par les intentions. Alors qu'il est possible (mais irrationnel) d'avoir des intentions incompatibles, il est *im*possible d'avoir des intentions incompatibles si ces dernières sont comprises comme des désirs dominants.

La seconde caractéristique de réduction posait que, avoir l'intention de faire X, c'est avoir un désir dominant de faire X ainsi que la croyance que je vais faire X en raison de ce désir. Clairement, cette deuxième tentative de réduction ne résiste pas à la dernière objection présentée dans le paragraphe précédent, où j'ai souligné que la relation d'incompatibilité imposée par le désir dominant n'avait rien à voir avec la relation d'incompatibilité imposée par l'intention. Aussi, si l'on définit l'intention conformément à cette deuxième tentative de réduction, on ne comprend pas que l'intention puisse mettre un terme à la délibération. Et comme les deux autres caractéristiques des intentions dépendent de la possibilité de mettre fin à la délibération, cette deuxième tentative de réduction n'explique pas davantage les deux autres caractéristiques des intentions. Les intentions sont donc bel et bien des états mentaux à part entière. Nous sommes justifiés de mettre de l'avant une architecture cognitive qui fasse une place importante aux intentions. Le reste du chapitre a tenté de définir cette architecture.

L'un des éléments centraux de l'architecture cognitive de Bratman *et al.* (1988) est le raisonneur procédant des fins aux moyens. Je me suis donc appliqué, au cours de ce chapitre, à définir ce raisonneur.

Le problème du raisonneur est de trouver (au moins) une séquence d'opérateurs qui va lui permettre de passer du modèle du monde actuel à un modèle du monde où le but est atteint. Donc, le problème du raisonneur est défini par trois inputs : un but, un modèle du monde actuel et une « banque » d'opérateurs. Le raisonneur résout son problème en adoptant la stratégie du GPS. Voici comment le GPS résout le problème du raisonneur.

À l'aide d'un démonstrateur de théorème (theorem-prover), le GPS essaie d'abord de démontrer l'intention ou but à partir du modèle du monde actuel. S'il peut démontrer cette intention à partir du modèle du monde actuel, le but est atteint dans ce monde. Le raisonneur n'a donc aucun opérateur à trouver. S'il n'est pas capable de démontrer le but à partir du modèle du monde actuel, il y aura un écart ou une différence entre le but et la conclusion à laquelle est parvenu le démonstrateur de théorème. L'objectif du raisonneur consiste alors à trouver des opérateurs dont les effets réduisent cette différence[24]. La prochaine étape consiste alors à appliquer l'opérateur, mais cela se fait à condition que les préconditions de l'opérateur

[24] On se rappelle qu'un opérateur est défini, d'une part, par les *effets* qu'il produit dans le monde et, d'autre part, par sa précondition, c'est-à-dire ses conditions d'applicabilité.

puissent être démontrées à partir du modèle du monde actuel. Si le démonstrateur de théorème peut démontrer les préconditions d'un opérateur, celui-ci est appliqué. Le raisonneur change alors (off-line) le modèle du monde conformément aux effets de l'opérateur. Le démonstrateur de théorème tente alors de démontrer le but à partir de ce nouveau modèle du monde. Si le but peut être démontré, le raisonneur a résolu son problème : il a trouvé un opérateur qui lui permet de passer du modèle du monde actuel à un modèle du monde où le but est atteint. Et si le but ne peut pas être démontré à partir du nouveau modèle du monde, l'opérateur répète les étapes qui viennent d'être décrite : il cherche de nouveaux opérateurs qui peuvent réduire la différence entre le but et le nouveau modèle du monde, il cherche à appliquer les opérateurs qui peuvent réduire cette différence, et il tente de démontrer le but à partir du modèle du monde qui résulte de l'application de ces nouveaux opérateurs.

Aussi, il faut dire que, lorsqu'un opérateur peut réduire la différence entre le but et le modèle du monde et que les préconditions de cet opérateur ne peuvent pas être démontrées à partir du modèle du monde, cet opérateur inapplicable n'est pas éliminé pour autant. Le raisonneur tente alors de trouver des opérateurs qui pourront réduire la différence qui existe entre les préconditions de cet opérateur et le modèle du monde. Les préconditions de ces opérateurs constituent alors des sous-buts. Et si ces opérateurs ne peuvent pas être appliqués étant donné que leurs préconditions ne peuvent pas être démontrées à partir du modèle du monde, le raisonneur tentera de trouver des opérateurs qui pourront réduire la différence entre le modèle du monde et les préconditions de ces opérateurs. Les préconditions de ces opérateurs seront alors des (sous-)sous-buts. Ce cycle peut se répéter, mais le raisonneur possède des heuristiques qui lui permettent de stopper sa recherche lorsque celle-ci devient trop compliquée ou sans espoir.

Après avoir trouvé un opérateur où une séquence d'opérateurs qui conduit à un modèle du monde où le but est atteint, le raisonneur envoie cet opérateur ou cette séquence d'opérateurs vers le filtre de compatibilité. Le filtre de compatibilité cherche alors à savoir si l'opérateur ou la séquence d'opérateurs est compatible avec les plans préalables de l'agent ainsi qu'avec ses croyances. Si l'opérateur est compatible, il est envoyé vers le processus de délibération, où il sera comparé à d'autres options (d'autres opérateurs compatibles). L'opérateur ou

l'option qui sera sélectionné par le processus de délibération ira ensuite composer le plan de l'agent.

Mais lorsque l'option est jugée incompatible avec le plan de l'agent et avec ce que l'agent croit, l'option n'est pas éliminée pour autant. Elle est envoyée vers le mécanisme de dérogation au filtre. Dans cette composante de l'architecture, il s'agit de voir si, malgré son incompatibilité, l'option vaut la peine d'être reconsidérée. S'il s'avère que ce n'est pas le cas, l'option est définitivement éliminée. Si l'option mérite d'être reconsidérée, cette option ainsi que les composantes du plan qui sont incompatibles avec elle sont envoyées vers le processus de délibération où un choix entre ces options sera fait. L'option sélectionnée ira ensuite alimenter le plan de l'agent.

Le raisonneur n'est pas le seul à émettre des options. L'analyseur d'opportunités émet aussi des options. L'analyseur d'opportunités, qui est un système de production, émet une option à chaque fois qu'un événement se présente comme une opportunité pour satisfaire un désir. L'option émise par l'analyseur d'opportunités suit le même chemin que les options du raisonneur : elle doit d'abord passer à travers le processus de filtration et, si elle passe avec succès à travers ce processus, elle doit ensuite être évaluée par le processus de délibération, qui se charge d'envoyer l'option sélectionnée vers le plan de l'agent.

Cette architecture cognitive permet d'établir des plans dont les intentions « descendent » jusqu'au niveau des habiletés. Les habiletés sont prises en charge par les perceptions d'affordances, et, comme on l'a vu au chapitre 1, ce sont les intentions du niveau inférieur qui contrôlent le système perceptuel (les perceptions d'affordances). Dès qu'une intention contrôle le système perceptuel, cette intention cesse d'alimenter le raisonneur pour que celui-ci lui trouve un moyen pour être réalisée. Ainsi, si l'on pouvait montrer que des intentions très générales (situées très haut dans la hiérarchie) contrôlent le système perceptuel, ces intentions très générales n'alimenteraient pas le raisonneur pour que celui-ci leur trouve un moyen pour être réalisées. Donc, cette architecture peut rendre compte d'un agent qui a un plan très peu détaillé. Autrement dit, cette architecture est compatible avec un agent réactif.

Mais qu'est-ce qui nous dit que cette architecture n'est pas en mesure de fournir un agent planificateur qui soit un expert? J'ai montré que, compte tenu du temps requis pour faire fonctionner le processus de délibération et le raisonneur procédant des fins aux moyens, il est

plausible de penser que cette architecture ne soit pas en mesure de gérer les événements inattendus au sein d'un environnement fortement dynamique; dans un environnement fortement dynamique, le temps de réaction qui est imposé par un événement inattendu est trop court pour que ces deux processus puissent gérer cet événement avec succès.

CHAPITRE III

LA PERCEPTION DES AFFORDANCES

Le modèle de l'action qui a été présenté au chapitre 1 repose en grande partie sur la possibilité de percevoir des affordances. Or, jusqu'à présent, la notion de perception d'affordance a seulement été esquissée. Compte tenu de l'importance de cette notion, il faut se donner la peine de l'approfondir. L'objectif du présent chapitre est précisément de définir la notion de perception d'affordance. Cependant, avant de déterminer ce qui caractérise la perception d'une affordance, il faudrait s'entendre sur ce qui caractérise les affordances. C'est sur ce dernier point que je vais m'attarder dans la première partie du chapitre 3. Dans la seconde partie, je vais me pencher sur la perception des affordances.

3.1 Les affordances

Les philosophes et psychologues qui se réclament du courant écologiste sont partagés entre deux façons de voir les affordances. Pour certains (Chemero, 2003; Stoffregen, 2003; Warren, 1984; Sanders, 1997), les affordances sont des relations entre l'agent et son environnement; pour d'autres (Greeno, 1994; Heft, 2001; Lombardo, 1987; Reed, 1996; Turvey, 1992; Scarantino, 2003), ce sont des propriétés de l'environnement. Dans les prochaines sections, je vais présenter la position d'un représentant majeur de chacune de ces deux façons de voir les affordances. Ces deux positions sont la position Stoffregen, parue dans son article *Affordances as properties of the animal-environment system* (2003), et la position de Turvey, parue dans son article *Affordances and prospective control : An outline of the ontology* (1992). Alors que Turvey considère les affordances comme des propriétés de l'environnement, Stoffregen croit que ce sont des relations entre l'agent et son

environnement. Comme on va le voir, je penche pour le premier terme de l'alternative; je crois que les affordances sont des propriétés de l'environnement.

Mais avant de présenter ces deux positions, déterminons ce qu'elles ont en commun.

À la suite de Gibson (1979/1986), tous s'entendent pour dire que l'affordance d'un objet prescrit à l'agent ce qu'il peut faire avec cet objet. Par exemple, si une chaise est dans mon environnement immédiat, la possibilité de m'asseoir m'est donnée; la chaise « afforde » l'action de s'asseoir. De même, si je suis sur un plancher, la possibilité d'être supporté m'est donnée; le plancher « afforde » le support. Les affordances sont donc des possibilités d'action ou possibilités behaviorales.

Aussi, ces deux positions s'entendent pour dire que les possibilités d'action ou affordances résultent d'une combinaison invariante de propriétés de surface et de substance considérées par rapport à l'agent. Il n'est pas difficile de se convaincre du fait qu'une combinaison invariante de propriétés de surface et de substance présente une affordance. Prenons l'exemple d'une chaise. Celle-ci « afforde » ou « offre » l'action de s'asseoir. Pour qu'elle « offre » cette action, il est nécessaire et suffisant que la chaise réunisse les cinq propriétés de surface et de substance suivantes : 1) sa surface est assez horizontale, 2) plate, 3) suffisamment étendue, 4) à peu près à hauteur des genoux, et 5) la substance de l'objet est suffisamment rigide. Si ces cinq propriétés de surface et de substance sont réunies dans un même objet, et seulement si elles le sont, la possibilité de s'asseoir est donnée par cet objet. Cependant, comme je l'ai dit au début du paragraphe, si les propriétés de surface et de substance sont prises en soi ou absolument, c'est-à-dire indépendamment de l'agent, la combinaison de ces propriétés ne présente pas l'affordance « assieds-toi ». Par exemple, une chaise « n'afforde » pas l'action de s'asseoir si elle est considérée relativement à un bébé. Pour qu'une chaise « afforde » cette action, la hauteur de la chaise doit être considéré relativement à la hauteur des jambes de l'agent; le ratio hauteur des jambes/hauteur de la chaise doit être à l'intérieur de certaines bornes.

Comme je l'ai dit, les deux positions annoncées plus haut reconnaissent que les affordances résultent de la combinaison invariante de propriétés de substance et de surface considérées par rapport à l'agent. Cependant, elles se distinguent à propos de l'endroit où situer l'affordance. Tandis que l'une de ces positions, celle de Stoffregen (2003), la situe

dans la relation entre les propriétés de l'environnement et l'agent, l'autre position, celle de Turvey (1992), la situe du côté de l'environnement. Plus précisément, chez Turvey, l'affordance est une propriété dispositionnelle de l'environnement : quand un objet de l'environnement réunit certaines propriétés de substance et de surface et que ces propriétés sont considérées relativement à l'agent, cet objet acquiert une disposition (une affordance). Je vais maintenant décrire la position de Turvey. Ensuite, je vais décrire la position de Stoffregen.

3.1.1. La position de Turvey (1992)

Les affordances sont des possibilités d'action. Donc, en tant que possibilités, elles sont des potentialités. À partir de ce constat, Turvey cherche dans le monde des propriétés qui se présentent comme des potentialités. Les dispositions s'offrent alors à lui comme un candidat idéal. Les dispositions, en effet, impliquent des événements qui ne sont pas actualisés. Par exemple, la disposition du sel à être soluble implique un événement non actualisé, la dissolution. Dans ce qui suit, je vais tâcher de déterminer les caractéristiques générales des dispositions et, ensuite, je vais tâcher de déterminer la forme particulière qui doit être donnée à ces caractéristiques lorsque les dispositions prennent la forme d'affordances. Après avoir présenté ces caractéristiques des affordances, je présenterai la définition formelle de Turvey.

Turvey attribue trois caractéristiques aux dispositions.

Premièrement, les dispositions viennent en paires. Par exemple, une chose X (par ex., la lumière) possède la disposition p à être réfractée et cette disposition n'est possible que s'il existe des substances réfractantes, que s'il existe une chose Z qui possède une disposition q à réfracter la lumière. Autrement dit, si une substance réfractante n'était pas en soi une chose possible, nous ne pourrions pas parler d'une substance qui a la disposition à être réfractée. On dit de la disposition q (la substance qui a la disposition de réfracter) qu'elle est le *complément* de la disposition p (la substance qui a la disposition à être réfractée)[25].

Deuxièmement, la disposition à faire X précède son actualisation ou, autrement dit, précède la réalisation de X. Quand je dis que la disposition de faire X précède son

[25] Si, au lieu de porter notre attention sur la disposition p, nous la portions sur la disposition q, l'inverse serait vrai. La disposition p serait le complément de la disposition q.

actualisation, je dis en fait que la disposition de faire X existe indépendamment du fait qu'elle soit ou non actualisée à travers X. Ainsi, le cristal possède la disposition à réfracter la lumière indépendamment du fait qu'un jet de lumière le traverse effectivement et qu'il (le cristal) réfracte effectivement ce jet de lumière.

Troisièmement, les dispositions ne manquent jamais de s'actualiser lorsque les circonstances appropriées se présentent[26]. Les circonstances permettant d'actualiser une disposition se ramènent à la présence du complément. Ainsi, si une chose X (par ex., la lumière) dotée d'une disposition p (être réfracté) est jointe à une chose Z (un cristal) dotée d'un complément q (la disposition à réfracter la lumière), la lumière va être réfractée. En fait, il serait plus juste de dire que, à travers une propriété r (c.-à-d. à travers la réfraction), les dispositions p et q vont être actualisées simultanément. Turvey souligne que, pour pouvoir être actualisées à travers la propriété r, les dispositions p et q doivent être filtrées par une fonction qu'il nomme « fonction de juxtaposition ». L'intervention de cette fonction de juxtaposition est nécessaire, car la chose X et la chose Z possèdent toutes deux une multitude de dispositions, si bien que, sans cette filtration (c.-à-d. l'exclusion des dispositions non pertinentes et la sélection des dispositions qui se complémentent), une multitude de dispositions pourraient s'actualiser. « In truth, *X* and *Z* have multiple dispositions – *m* and *n*, respectively. To actualise W_{pq}, the juxtaposition function *j* must be such as to filter *p* and *q* from the array of $m \times n$ dispositions possessed by *X* and *Z*. » (Turvey, 1992, p. 179)

À la lumière de ce qui vient d'être dit sur les dispositions, précisons maintenant les caractéristiques des affordances.

Premièrement, les affordances ont besoin d'un complément. Le complément de l'affordance est nommé par Turvey « effectivité » (effectivity). L'effectivité est la disposition d'un agent à faire une certaine action (tandis que l'affordance est une disposition de l'objet à permettre cette même action). Et il est clair que l'affordance a besoin de son complément. Prenons une chaise. Si aucun agent n'avait la disposition à s'asseoir (l'effectivité), ou, encore

---

[26] Il faut voir que cette dernière caractéristique pourrait nécessiter l'ajout d'une clause *ceteris paribus*. Pour se faire une idée à ce sujet, il faudrait interroger le scientifique et lui demander si les lois physiques qui concernent les dispositions dépendent du contexte. En outre, il est possible que la plupart des lois physiques, et pas seulement celles qui concernent les dispositions, requièrent l'ajout d'une clause *ceteris paribus*.

plus radicalement, si la disposition à s'asseoir était en soi impossible, alors les chaises « n'afforderaient » pas l'action de s'asseoir. Ou, de manière plus particulière, si un agent n'avait pas la disposition à s'asseoir, les chaises « n'afforderaient » pas pour cet agent l'action de s'asseoir. Cette caractérisation de l'affordance, c'est-à-dire le fait qu'elle a besoin d'un complément pour exister, permet de définir les affordances comme des propriétés relationnelles. Lorsque nous aborderons la question de la *perception* des affordances, il faudra tenir compte de cette relation à l'effectivité. Plus précisément, la perception d'une affordance devra intégrer cette relation à l'effectivité.

Selon la deuxième caractéristique énoncée plus haut, la disposition précède son actualisation. Donc, l'affordance précède aussi son actualisation. Autrement dit, même lorsque l'agent n'est pas dans les parages et qu'aucune action n'est donc actualisée, l'affordance existe.

Troisièmement, à la lumière de la dernière caractéristique dégagée ci-dessus, il faut poser que l'affordance ne manque jamais de s'actualiser lorsqu'elle est unie à son complément, c'est-à-dire l'effectivité qui lui correspond. C'est la fonction de juxtaposition qui unit l'affordance et l'effectivité. Elle rend cette union possible en filtrant l'affordance et l'effectivité. En effet, l'objet possède une multitude d'affordances et l'animal possède aussi une multitude d'effectivités correspondantes, si bien que, sans cette fonction de juxtaposition, une multitude d'affordances et d'effectivités correspondantes seraient actualisées à chaque fois que l'agent entre en contact avec cet objet. À la lumière de ce qui a été dit au chapitre 1, nous pouvons poser que c'est l'intention qui filtre l'affordance et l'effectivité. En effet, au chapitre 1, nous avons vu que l'intention fait en sorte que le système perceptuel sélectionne une affordance plutôt que les autres et qu'elle fait en sorte que l'agent agit conformément à cette affordance. On doit souligner ici une différence importante entre les dispositions purement physiques (comme la disposition de la lumière à être réfractée) et les dispositions qui font intervenir des affordances. Dans une disposition purement physique, la disposition et son complément sont automatiquement sélectionnés ou filtrés sur la base de mécanismes physiques; pour les affordances, la sélection de l'affordance et de son complément se fait sur la base d'un état cognitif, à savoir l'intention.

Ce qui précède semble suggérer que les animaux non humains possèdent des intentions. Pas nécessairement. Pour l'animal (humain et non humain), la filtration d'une affordance consiste dans le fait de faire attention à cette affordance plutôt qu'aux autres. Donc, chez l'humain, l'intention filtre une affordance parce qu'elle fait en sorte que l'agent fait attention à cette affordance plutôt qu'aux autres. Mais il est possible que ce qui fait qu'un animal non humain prête attention à une affordance plutôt qu'aux autres, ce soit, non pas une intention, mais un autre type d'état biologique. Ce sera donc cet état biologique (qui ne serait pas une intention) qui, chez l'animal non humain, filtrera l'affordance.

Après avoir attribué ces caractéristiques de base aux affordances, Turvey énonce une définition formelle des affordances :

> Let W_{pq} (e.g., a person-climbing-stairs system) = $j(X_p, Z_q)$[27] be composed of different things Z (person) and X (stairs). Let p be a property of X and q a property of Z. Then p is said to be an affordance of X and q the effectivity of Z (i.e., the complement of p), if and only if there is a third property r such that
> (i) $W_{pq} = j(X_p, Z_q)$ possesses r
> (ii) $W_{pq} = j(X_p, Z_q)$ possesses neither q nor p
> (iii) Neither Z nor X possesses r (Turvey, 1992, p. 180)

La condition (i) nous dit que, si un agent ayant la disposition de monter est uni (via la fonction de juxtaposition) à un objet qui possède la disposition à être monté (c.-à-d. une surface inclinée composée de marches ayant une hauteur adéquate et une surface horizontale suffisante), la montée de l'escalier (r) est actualisée au sein du système formé par cet agent et cet objet. La condition (iii) nous dit que, pris isolément, l'agent et l'escalier ne possèdent pas les propriétés qui sont propres à l'actualisation. « The actualizing engenders new locomotory dynamics that are not present in standing and walking, and new reactive force from the supporting surfaces that are not present in the absence of being climbed on. » (Turvey, 1992, p. 180) La condition (ii) nous dit que l'affordance et l'effectivité ne se trouvent pas dans le système unifié, où se produit l'actualisation. Afin de bien faire comprendre cette condition, Turvey réfère à une disposition physique bien connue : « [t]he disposition p to be soluble rests with the fact that it is a lattice of electrically charged ions bound by an electrical attraction between opposite charge that can be eliminated by a liquid with a high dielectric

[27] La lettre « j » réfère à la fonction de juxtaposition.

constant. The salt-dissolved-in-water system lacks the attraction between ions; it does not possess p. » (Turvey, 1992, p. 181)

3.1.2 La position de Stoffregen (2003)

La position de Stoffregen est émergentiste. L'émergentisme pose que le tout comprend des propriétés dont sont dépourvues les parties. Le tout se réduit à la relation entre les parties. Alors, selon les émergentistes, la relation entre les parties possède des propriétés dont sont dépourvues les parties. Ainsi, le triangle, qui se ramène à une relation entre des droites, possède des propriétés dont sont dépourvues les droites considérées isolément. Relativement aux affordances, la position de Stoffregen est celle-ci : l'affordance émerge de la relation entre les propriétés de l'environnement et les propriétés de l'agent. Donc, chez Stoffregen, l'affordance ne se trouve ni chez l'agent ni dans l'environnement, mais dans la relation entre ces derniers; c'est le système environnement-agent qui possède des affordances. « Because they arise from relations between animal and environment, affordances are emergent properties of the animal-environment system. » (Stoffregen, 2003, p. 124) Ainsi, si l'on considère de nouveau l'affordance présentée par un escalier, cette affordance se ramène chez Stoffregen à la relation entre les marches de l'escalier et la hauteur des jambes de l'agent. À part Turvey, je crois que Stoffregen est le seul à avoir donné une définition formelle des affordances. Conformément à la position émergentiste qu'il défend, voici comment Stoffregen définit les affordances :

> Let W_{pq} (e.g., a person-climbing-stairs system) = (X_p, Z_q) be composed of different things Z (e.g., person) and X (e.g., stairs)
>
> Let p be a property of X and q be a property of Z.
>
> The relation between p and q, p/q, defines a higher order property (i.e., a property of the animal-environment system), h.
>
> Then h is said to be an affordance of W_{pq} if and only if
> (i) $W_{pq} = (X_p, Z_q)$possesses$h$
> (ii) Neither Z nor X possesses h (Stoffregen, 2003, p. 123)

La condition (i) nous dit que l'affordance, h ou p/q, relève du système agent-environnement. La condition (ii) nous dit que l'affordance (h) ne se trouve ni dans l'objet ni chez l'agent pris isolément. Ceci est conforme à la conception émergentiste, qui pose que les

propriétés de la relation (le tout) ne se trouvent pas dans les parties. Aussi, il faut noter que les propriétés *p* et *q* n'ont pas besoin d'être jointes par une fonction de juxtaposition. Cela tient au fait que la relation entre p et q, c'est-à-dire l'affordance h, ne déclenche pas automatiquement l'actualisation de l'affordance.

J'ai dit ci-dessus que je n'adoptais pas la définition de Stoffregen. Mon argument est assez direct.

Les affordances sont des *possibilités* d'action ou, autrement dit, des potentialités. Or, la propriété qui émerge de la relation entre plusieurs parties est une propriété *actuelle*. La propriété qui émerge de la relation entre les propriétés de l'objet et les propriétés de l'agent ne peut donc pas être une affordance. Cette objection ne touche évidemment pas la définition de Turvey, puisque sa définition définit les affordances comme des dispositions et que les dispositions sont des potentialités.

Dans son article de 2003, Stoffregen a critiqué la définition de Turvey. Donc, si je veux adopter la définition de Turvey, c'est-à-dire définir les affordances comme des dispositions, je dois répondre aux objections de Stoffregen.

3.1.3. Les objections de Stoffregen (2003)

Stoffregen présente deux objections à la position de Turvey. La première objection concerne la troisième caractéristique que Turvey attribue aux dispositions. Selon cette caractéristique, lorsqu'une disposition est unie à la disposition qui la complémente, l'actualisation s'ensuit automatiquement. A priori, cette caractéristique pose problème car, lorsque l'agent est en contact avec un objet, cet objet peut « afforder » une multitude d'actions et l'agent n'actualise qu'une seule de ces actions. Or, si la troisième caractéristique est valide, l'agent devrait actualiser toutes ces actions. De plus, ce problème est encore plus criant si l'on considère que, quelques fois, l'objet peut « afforder » des actions incompatibles entre elles. Turvey résout ce problème en introduisant un filtre (une fonction de juxtaposition). Comme on l'a vu, ce filtre sélectionne un seul couple affordance-effectivité et c'est ce couple particulier qui est actualisé. Stoffregen est conscient de la solution de Turvey. Cependant, Stoffregen remarque que l'introduction de ce filtre pose un autre problème :

Turvey n'a pas défini ce filtre, si bien que sa définition des affordances comme propriétés dispositionnelles est incomplète.

> The need for a filter renders Turvey's (1992) formal definition fatally incomplete. The filter is not defined in the definition of *affordances*, and for this reason the definition cannot account for the actualisation of any given affordance when multiple affordances exist. Because multiple affordances always exist, Turvey's definition is incomplete for any real case. (Stoffregen, 2003, p. 120)

Nous pouvons répondre à l'objection de Stoffregen. Comme je l'ai indiqué ci-dessus, le filtre que Turvey n'a pas pris la peine de définir a été défini par nous. Ci-dessus, nous avons vu que ce filtre correspondait à l'intention de l'agent. C'est cette dernière qui explique comment un couple affordance-effectivité est sélectionné au détriment des autres couples affordance-effectivité. Le filtre a donc été défini par nous. La définition de Turvey n'est pas incomplète.

La deuxième objection de Stoffregen concerne la possibilité de percevoir *directement* l'affordance. Stoffregen nous dit que, lorsque les affordances sont comprises comme le fait Turvey, c'est-à-dire comme des propriétés de l'environnement (comme des dispositions de l'environnement), les affordances doivent être perçues de manière indirecte, ce qui, dans une perspective écologique, équivaut à une objection sérieuse[28].

Pour bien faire comprendre l'objection de Stoffregen, je dois donner un aperçu de la théorie de la perception des affordances de Turvey. Cet aperçu pourrait être difficile d'accès pour celui qui n'est pas déjà familier avec la théorie de la perception des affordances de Turvey. Je tiens à souligner que les propos qui seront tenus ici seront explicités dans la deuxième partie du chapitre 3, où je rends compte de la possibilité de percevoir des affordances.

Chez Turvey, l'affordance est une disposition et, donc, à ce titre, une propriété non observable. Comment rend-t-il compte de la possibilité de percevoir ce genre de propriété? Il en rend compte en terme informationnel. Percevoir une affordance, c'est acquérir l'information qui porte sur cette propriété. Pour acquérir l'information qui porte sur une

---

[28] En effet, dans une perspective écologique, la perception des affordances est directe, non médiatisée par des inférences.

propriété X, il faut saisir une propriété Y qui est corrélée nomologiquement à la propriété X. En effet, dans ces circonstances, la propriété Y donne de l'information sur la propriété X. Nous pouvons également dire que la propriété Y spécifie la propriété X. La théorie de la perception des affordances de Turvey pose que les propriétés visibles de l'objet spécifient l'affordance (la propriété non observable). Ainsi, les différentes propriétés visibles d'une chaise (sa hauteur, sa rigidité, etc.) sont corrélées nomologiquement à l'affordance, la possibilité de s'asseoir. Ensuite, Turvey pose que la structure lumineuse réfléchie par la chaise spécifie les propriétés visibles (ou est corrélée nomologiquement à ces dernières)[29]. Donc, par transitivité, la structure lumineuse spécifie l'affordance. En saisissant cette structure lumineuse, nous possédons l'information qui porte sur l'affordance; nous percevons alors l'affordance. C'est ainsi que Turvey comprend la possibilité de percevoir des affordances.

Stoffregen critique cette façon de spécifier les affordances. Selon lui, cette manière de concevoir la spécification de l'affordance ne tient pas compte de la complémentarité de l'affordance avec l'effectivité. Comme on l'a vu, une affordance se présente seulement lorsqu'elle est mise en relation avec l'effectivité de l'agent. Donc, une théorie de la perception des affordances doit intégrer en son sein une spécification de cette relation à l'effectivité. Or, la théorie de Turvey ne le fait pas. On y trouve une spécification de l'affordance par les propriétés occurrentes et une spécification des propriétés occurrentes par la structure lumineuse, mais aucune spécification de la relation à l'effectivité. Or, comme la relation à l'effectivité est essentielle à l'affordance et à sa perception, Turvey doit insérer cette relation dans le processus perceptuel. Comment peut-il s'y prendre? Comme cette relation n'est pas spécifiée par la structure optique et comme la perception directe suppose la spécification par la structure optique, Turvey ne peut intégrer cette relation à l'effectivité que de manière indirecte; l'esprit doit y mettre du sien. La théorie de la perception des affordances de Turvey comprend donc la perception des affordances comme une perception indirecte.

---

[29] Turvey (1992) ne mentionne que la structure lumineuse comme porteuse d'information sur les propriétés occurrentes ou actuelles (la hauteur de la chaise, l'étendue de sa surface, etc.). Or, la perspective écologique tient aussi compte des autres modalités sensibles pour percevoir les affordances. Pour simplifier mes descriptions, cependant, je me limiterai la modalité visuelle.

Il est possible de répondre à l'objection de Stoffregen en augmentant la théorie de Turvey de manière à lui incorporer cette relation à l'effectivité. Il suffit de poser que, lors de la spécification des propriétés occurrentes (les propriétés de surface et de substance) par la structure lumineuse, ces propriétés occurrentes sont mises à l'échelle (scaled) de l'effectivité, c'est-à-dire, plus précisément, mises à l'échelle des propriétés occurrentes qui sont à la base de l'effectivité. De cette manière, la structure lumineuse spécifie non seulement les propriétés occurrentes, mais aussi la relation de ces propriétés à l'effectivité. La perception des affordances devient donc une perception directe. Je vais revenir plus tard sur cette question de la mise à l'échelle (scaling).

3.2. La perception des affordances

Selon Gibson (1979/1986), pour qu'un système perceptuel perçoive une affordance, il suffit qu'il ait en sa possession l'information portant sur l'affordance. Selon Gibson, cette information est portée par la structure optique. Donc, pour qu'il y ait perception d'une affordance, il suffit que le système perceptuel saisisse ou capte la structure optique (le signal) qui porte l'information sur l'affordance. Pour s'assurer de la plausibilité de la position de Gibson sur la perception des affordances, il faudrait donc s'assurer de deux choses : d'une part, que la structure optique puisse transmettre l'information à propos de l'affordance; d'autre part, que le système perceptuel soit en mesure de saisir ou d'extraire la structure optique qui porte l'information sur l'affordance. Dans le chapitre 4, je vais présenter un mécanisme perceptuel qui explique comment la saisie ou extraction de cette information est possible. Dans la section 3.2.1, je vais m'assurer de la capacité de la structure optique à pouvoir transmettre l'information à propos de l'affordance. Dans la section 3.2.2, je vais déterminer ce qui caractérise la structure optique lorsqu'elle porte de l'information sur une affordance. Nous allons voir que le signal optique portant de l'information sur l'affordance se caractérise comme un *invariant d'invariants*. Après avoir analysé la structure optique, je vais me pencher sur un autre aspect de la perception des affordances. Ci-dessus, nous avons vu que les affordances étaient des propriétés relationnelles; qu'elles étaient mises en relation avec l'effectivité de l'agent. Dans la section 3.2.3, je vais m'assurer que la perception des affordances intègre cette relation à l'effectivité.

3.2.1 La structure optique peut-elle transmettre de l'information sur l'affordance?

Cette section se divise en deux parties. Tout d'abord, je vais présenter la théorie de l'information telle que présentée par Dretske dans son texte de 1981, *Knowledge and the flow of information*. Ma lecture du texte de Dretske va me permettre de déterminer les conditions qu'un signal doit satisfaire pour pouvoir transmettre de l'information à propos d'une propriété. Ensuite, je vais voir si la structure optique satisfait ces conditions lorsque les propriétés sont des affordances. Selon que la structure optique satisfait ou non ces conditions, je pourrai me prononcer sur la capacité de la structure optique à transmettre l'information à propos d'une affordance. Commençons donc l'analyse du texte de Dretske.

3.2.1.1 La théorie de Dretske

L'occurrence d'un état de choses génère avec elle une quantité d'information. La quantité d'information générée par l'occurrence d'un état de choses correspond à la réduction de l'incertitude ou, autrement dit, à l'élimination des possibilités. Plus la réduction de l'incertitude est grande, plus la quantité d'information est grande. Ainsi, si plusieurs états de choses peuvent se produire, l'occurrence de l'un de ces états de choses génère plus d'information que si les états de choses pouvant se produire avaient été moins nombreux. Cette définition de la quantité d'information — quantité d'information = réduction de l'incertitude — est parfaitement intuitive. Ainsi, si je connais l'identité du prochain événement, l'occurrence de l'événement en question m'apprendra beaucoup moins de choses (me donnera moins d'information) que si je ne connaissais pas son identité. Habituellement, on calcule la réduction de l'incertitude en termes de décisions binaires, que l'on qualifie « bits ».

Ceci étant dit, revenons maintenant à la question qui nous occupe : À quelle(s) condition(s) le signal r transmet-il l'information que s est P? Dretske (1981) énonce trois conditions. La première condition est que la quantité d'information transmise par le signal r

doit être *égale* à la quantité d'information générée par l'occurrence de l'état de choses *s* est *P*[30].

Ce qui motive Dretske à adopter cette première condition, c'est que le rejet de cette condition implique la violation d'un principe que Dretske nomme « principe Xérox ». Autrement dit, lorsque l'on accepte l'idée que la quantité d'information transmise par le signal peut être *moins* grande que la quantité d'information générée à la source *s*, on viole le principe Xérox. Je vais maintenant définir le principe Xérox et je vais ensuite montrer qu'il est vrai que le rejet de cette première condition implique une violation de ce principe.

Le principe Xérox pose que « if *A* carries the information that *B*, and *B* carries the information that *C*, then *A* carries the information that *C*. » (Dretske, 1981, p. 57) Ce principe a une base intuitive. Il reflète l'idée intuitive que si je peux apprendre de *A* que *B* et de *B* que *C*, alors je peux apprendre de *A* que *C*. Ce principe est en jeu dans les phénomènes qui manifestent un flux d'information. Ainsi, le principe Xérox est en jeu à chaque fois que les ondes acoustiques qui émanent de mon haut-parleur portent de l'information sur ce qui se passe à la station de radio émettrice. Tout d'abord, on constate l'existence de plusieurs liens de communication. Les ondes acoustiques portent de l'information sur ce qui se passe dans les circuits électriques du haut-parleur; les circuits électriques portent de l'information sur ce qui se passe dans les antennes réceptrices (la réception d'ondes électromagnétiques); les ondes électromagnétiques portent de l'information sur la façon dont vibre le microphone; et, enfin, les vibrations du microphone portent de l'information sur ce qui est dit par l'animateur. Maintenant, ce qui fait que, malgré ces liens de communication intermédiaires, les ondes acoustiques portent de l'information sur ce qui se dit dans la station de radio, c'est que chaque lien de communication porte de l'information sur le lien qui le précède. Autrement dit, c'est grâce à une itération du principe Xérox que les ondes acoustiques portent de l'information sur les paroles de l'animateur.

---

[30] Le signal transmet la même quantité d'information que la quantité d'information générée par l'état de choses (la cause du signal) si le signal renvoie de manière univoque à cet état de choses. La transmission d'information se caractérise par une perte d'information lorsque le signal peut être causé par des états de choses différents; autrement dit, lorsque le signal manifeste de l'équivoque ou de l'ambiguïté.

Une théorie de l'information ne peut pas mettre de côté le principe Xérox. De l'aveu même de Dretske, le principe Xérox est « a regulative principle, something inherent in and essential to the ordinary idea of information, something that any *theory* of information should preserve. » (Dretske, 1981, p. 57) Je vais maintenant montrer, m'inspirant toujours de Dretske (1981), que le rejet de la condition énoncée ci-dessus implique une violation de ce principe. Pour cette raison, nous devrons maintenir cette première condition.

L'objectif est donc de montrer que, si la quantité d'information portée par le signal est *moins* grande que la quantité d'information générée par la source, le principe Xérox est violé. Afin de se rendre la tâche plus ardue, supposons que l'information portée par le signal est à peine moins grande que la quantité d'information générée par l'occurrence de l'état de choses et supposons (à tort) que lorsque la perte d'information est aussi minime, la transmission d'information par le signal est un succès.

Considérons d'abord une chaîne de communication ayant deux liens de communication. *A* transmet l'information que *B* et *B* transmet l'information que *C*. Et supposons, conformément au paragraphe précédent, que chaque transmission se caractérise par une perte d'information et que cette perte d'information ne soit pas suffisante pour éliminer le message porté par le signal. Autrement dit, malgré la perte d'information, le signal *A* transmet l'information que *B*, et *B* transmet l'information que *C*. Cependant, ces pertes d'information s'additionnent lorsque l'on passe d'une transmission à l'autre. La quantité d'information du signal *A* à propos de *C* est moins grande que la quantité d'information du signal *B* à propos de *C*[31]. Donc, si l'on multiplie les liens de communication; si l'on suppose, par exemple, que C transmet l'information que *D*; que *D* transmet l'information que *E* ... et, finalement, que *Y* transmet l'information que *Z*, alors, compte tenu que la perte d'information s'additionne, l'information que le signal *A* porte sur le premier état de choses, *Z*, est immensément moins grande que l'information générée par *Z*. Or, il est clair que si *A* contient beaucoup moins d'information que l'information générée par *Z*, *A* ne peut pas porter l'information que *Z*. Donc, le principe Xérox est violé. Donc, pour que le principe Xérox soit sauvegardé, il faut qu'il n'y ait aucune perte d'information lors de la transmission d'information; il faut que la

[31] Pour une démonstration quantitative du fait que la perte d'information s'additionne lorsque l'on passe à une autre transmission d'information, voir Dretske (1981, note n° 22 du chapitre 2).

quantité d'information du signal soit aussi grande que la quantité d'information générée à la source.

Maintenant, montrons que cette première condition n'est pas suffisante. Autrement dit, montrons que le signal *r* peut très bien porter la même quantité d'information que l'occurrence « s est *P* » sans pour autant porter l'information que *s* est *P*.

L'occurrence « *s* est *Q* » peut générer la même quantité d'information que l'occurrence « *s* est *P* », si bien que, lorsque le signal r porte une quantité d'information qui est égale à la quantité d'information générée par l'occurrence « *s* est *P* »; le signal *r* peut porter l'information que *s* est *Q*. Il faut donc ajouter une condition pour comprendre la capacité de *r* à transmette l'information que *s* est *P*. La deuxième condition est : *s* est *P* (et non pas *Q*). Avec l'ajout de cette condition, le signal *r* ne peut pas transmettre l'information que *s* est *Q*. En effet, puisque l'état de chose « *s* est *Q* » n'est pas présent, l'information que *s* est *Q* serait une information fausse ou erronée, et l'erreur ou la fausseté ne sont pas des attributs du concept d'information (Dretske, 1981, p. 45)[32].

Même combinées, les deux conditions précédentes ne sont pas suffisantes pour que l'on puisse dire que le signal *r* porte l'information que *s* est *P*. Supposons que *s* soit à la fois carré et rouge (*s* est un carré rouge). Le symbole *P* représente la propriété « rouge »; le symbole *K*, la propriété « carré ». De plus, supposons que l'occurrence de ces deux propriétés génère la même quantité d'information, disons 3 bits. Un signal est alors transmis et celui-ci porte l'information que *s* est *K*. Ce signal porte 3 bits d'information (satisfaction de la première condition), et *s* est *P* (satisfaction de la deuxième condition). Pourtant, ce signal ne porte pas l'information que *s* est *P*. Donc, le problème ici, c'est que, bien que les deux conditions soient satisfaites, l'information portée par le signal *r* ne correspond pas à l'information qui est générée par l'occurrence de l'état de choses « *s* est *P* ». La troisième condition est donc :

---

[32] « [...] *false* information and *mis*-information are not kinds of information – any more than decoy ducks and rubber ducks are kinds of ducks. And to speak of certain information as being reliable is to speak redundantly. » (Dretske, 1981, p. 45)

la quantité d'information portée par le signal est la quantité d'information généré par l'occurrence « *s* est *P* »[33].

De l'aveu même de Dretske, cette troisième condition n'est pas très claire, car on ne voit pas vraiment comment une quantité d'information peut correspondre à une autre quantité d'information, si ce n'est par une comparaison numérique. On peut néanmoins conserver cette condition, car cette condition « is merely intended to mark, in a suggestive way, the need for something beyond (A) and (B) [les deux premières conditions] [...]. » (Dretske, 1981/1999, p. 64)

Donc, si ces trois conditions sont respectées, le signal *r* porte ou transmet l'information que *s* est *P*. Toute définition de la transmission d'information par un signal doit donc satisfaire ces trois conditions. Dretske définit ainsi la transmission d'information par un signal *r* : dire qu'un signal *r* porte l'information que *s* est *P*, c'est dire que la probabilité conditionnelle que *s* soit *P*, étant donné le signal *r*, est de 1[34].

Dretske (1981/1999) montre que cette définition satisfait les trois conditions énoncées ci-dessus.

Premièrement, comme cette définition pose que la probabilité conditionnelle de l'état de choses « *s* est *P* », étant donné le signal *r*, est de 1, le signal *r* doit renvoyer de manière univoque à l'état de choses « *s* est *P* »; il ne doit pas y avoir d'équivoque. Or, s'il n'y a pas d'équivoque entre le signal et l'état de chose, alors la quantité d'information portée par ce signal est égale à la quantité d'information générée par l'occurrence de l'état de choses. Cela découle de cette équation de la théorie de l'information de Shannon : la quantité d'information transmise par le signal = la quantité d'information à la source — équivoque. La première condition est donc satisfaite. Deuxièmement, comme la probabilité conditionnelle de l'état de choses « *s* est *P* », étant donné le signal *r*, est de *1*, l'état de choses « *s* est *P* » doit avoir lieu lorsque *r* survient. La deuxième condition — *s* est *P* — est donc satisfaite. Troisièmement, comme cette définition pose que la probabilité conditionnelle de

---

[33] Ma formulation de la troisième condition correspond presque mot à mot à celle de Dretske : « The quantity of information the signal carries about *s* is (or includes) that quantity generated by *s*'s being *F* (and not, say, by *s*'s being *G*). » (1981, p. 64)

[34] « A signal r carries the information that s is F = The conditional probability of s's being F, given r [...], is 1 [...]. » (Dretske, p. 65, 1981/1999)

l'état de choses « *s* est *P* », étant donné *r*, doit être de 1, cette définition exclut les situations (les autres états de choses) qui nous ont forcés à introduire la troisième condition. Cette définition suppose donc que la quantité d'information portée par le signal *r* est la bonne quantité d'information, c'est-à-dire la quantité d'information qui correspond à la quantité d'information générée par l'occurrence de l'état de choses « *s* est *P* ». La définition de Dretske satisfait donc la troisième condition.

Donc, on peut effectivement dire qu'un signal *r* transmet l'information que *s* est *P* si, et seulement si, la probabilité conditionnelle de l'état de choses « *s* est *P* », étant donné *r*, est de 1. On pourrait être tenté de donner une interprétation causale à cette définition de la transmission d'information par un signal. Plus précisément, on pourrait être tenté de dire qu'un signal *r* transmet l'information que *s* est *P* si, et seulement si, l'état de choses « s est P » est la cause unique du signal r. Il est vrai que, si l'état de choses « s est P » est la cause unique du signal *r*, le signal *r* transmet l'information que *s* est *P*. En effet, si l'état de choses « s est P » est la cause unique du signal r, alors la probabilité conditionnelle de l'état de choses « s est P », étant donné le signal r, est de 1. Cependant, ce n'est pas *seulement si* l'état de choses « s est P » est la cause unique du signal *r* que le signal *r* transmet l'information que *s* est *P*. En effet, le signal *r* peut transmettre l'information que *s* est *P* sans que le signal *r* et l'état de choses « s est P » soient dans une relation causale. Premièrement, la probabilité conditionnelle de l'état de choses « s est P », étant donné le signal r, peut être de 1 si le signal *r* et l'état de choses « s est P » sont causés par une cause commune. Lorsque c'est le cas, c'est-à-dire lorsque le signal *r* et l'état de choses « s est P » sont causés par une cause commune, l'état de choses « s est P » ne cause pas le signal *r*, et pourtant la probabilité conditionnelle de l'état de choses « s est P », étant donné le signal *r*, est toujours de 1; autrement dit, le signal *r* transmet l'information que *s* est *P*. Il existe une autre situation où le signal *r* transmet l'information que *s* est *P* sans que le signal *r* et l'état de choses « s est P » soit dans une relation causale. Cette situation fait référence à un monde non déterministe ou, autrement dit, à un monde *non* causal[35]. Dans un tel monde, l'état de choses « s est P » peut produire, à différentes occasions, des signaux différents. À une occasion, l'état de choses « s est P » produit le signal r_1 et, à une autre occasion, il produit le signal r_2. Et supposons que

[35] Pour se persuader qu'un monde non déterministe est équivalent à un monde non causal, voir Dretske (1981).

les signaux r_1 et r_2 ne peuvent pas être produits par plus d'un état de choses. Lorsque le signal r_1 se produit, il est certain que l'état de choses « s est P » s'est produit; idem pour le signal r_2. Autrement dit, la probabilité conditionnelle de l'état de choses « s est P », étant donné le signal r_1 (ou r_2), est de 1. Donc, le signal r_1 (ou r_2) transmet l'information que s est P. Pourtant, comme il n'y a pas de relation déterministe entre l'état de choses « s est P » et le signal r_1 (ou r_2), le signal r_1 (ou r_2) et l'état de choses « s est P » ne sont pas dans une relation causale.

Donc, la définition drestkienne de la transmission d'information par un signal *r* ne suppose pas nécessairement une relation causale entre l'état de choses « s est P » et le signal *r*. Mon objectif, maintenant, est de reformuler la définition de Dretske d'une façon qui soit plus brève. Comment, puisqu'il n'est pas légitime de parler d'une relation causale entre l'état de choses « s est P » et le signal *r*, peut-on caractériser la relation du signal r à l'état de choses « s est P »? Je vais poser qu'il s'agit, non pas d'une relation de causalité, mais d'une relation de corrélation. Si le signal r est *corrélé* à l'état de choses « s est P », et si cette corrélation est de 1, le signal *r* transmet l'information que *s* est *P*. Cette reformulation de la définition de Dretske permet de rendre compte du cas de figure causal. En effet, s'il y a une relation causale entre deux événement et si la cause est unique, ces deux événements (le signal r et l'état de choses « s est P ») sont corrélés. Cette reformulation permet également de rendre compte du premier cas de figure non causal, où le signal r et l'état de choses « s est P » sont causés par une cause commune. En effet, dans ce cas de figure, le signal r est corrélé à l'état de choses « s est P ». Cette reformulation, cependant, répond moins bien au deuxième cas de figure, celui qui fait intervenir un monde non causal. En effet, peut-on dire que le signal r_1 est corrélé à l'état de choses « s est P »? S'il faut identifier « covariation » et « corrélation », non. En effet, bien qu'une variation du signal r_1 soit nécessairement accompagnée d'une variation de l'état de choses « s est P », l'inverse n'est pas vrai : une variation de l'état de choses « s est P » n'est pas nécessairement accompagnée d'une variation du signal r_1; il est possible que ce soit r_2 qui varie. Donc, notre reformulation de la définition de Dretske n'est pas parfaite. Je vais néanmoins la conserver. Comme l'action se situe à une échelle de grandeur où il n'y a probablement pas de comportement non déterministe, ce défaut de notre définition de la transmission d'information par un signal *r* est sans grande conséquence.

On peut donc affirmer que la structure optique donne une information sur l'affordance si, et seulement si, la structure optique est corrélée nomologiquement[36] à l'affordance. Je vais maintenant montrer que la structure optique est effectivement corrélée nomologiquement à l'affordance.

3.2.1.2. La structure optique porte-t-elle l'information sur l'affordance?

La *physique écologique* soutient que les propriétés occurrentes (les propriétés de surface et de substance) sont corrélées nomologiquement aux affordances (Turvey *et al.*, 1981, p. 264). En effet, une variation des propriétés occurrentes est accompagnée d'une variation des affordances et inversement. Par exemple, si je varie les affordances, si je passe de l'affordance « assieds-toi » à l'affordance « frappe-moi », nous retrouverons une variation correspondante chez les propriétés occurrentes. Et l'inverse est aussi vrai : une variation des propriétés occurrentes est accompagnée par une variation des affordances. Donc, il est vrai que les propriétés occurrentes sont corrélées aux affordances. Et cette corrélation est nomologique, car elle est *toujours* valable, elle n'est pas accidentelle.

Et la structure optique est aussi corrélée nomologiquement aux propriétés occurrentes. C'est là une des thèses centrales de la psychologie écologique de Gibson. Gibson a effectivement analysé la structure optique et il a montré qu'une variation de la structure optique était accompagnée par une variation de la propriété occurrente. Ainsi, entre autres choses, il a montré qu'une structure optique bien particulière correspondait à la surface *arrière* d'un objet (Gibson, 1979/1986). Dans le présent travail, je ne reprendrai pas les analyses de Gibson; pour se convaincre de la plausibilité de la thèse de Gibson, je renvoie le lecteur au texte de Gibson, *The ecological approach to visual perception* (1979/1986). Cependant, afin d'asseoir la thèse de Gibson sur les bases les plus solides qui soient, je tâcherai de réfuter les voix qui se sont objectées à sa thèse. Je pense ici essentiellement au texte publié en 1981 par Fodor et Pylyshyn, *How direct is visual perception? Some reflections on Gibson's « ecological approach »*. C'est au cours du chapitre 5 que je tâcherai

[36] Le terme « nomologique » est introduit pour signifier que la corrélation est toujours valable; qu'elle n'est pas accidentelle. De plus, il faut se souvenir que, pour être compatible avec la définition de Dretske, cette corrélation doit être de 1.

de réfuter les arguments de cet article. Mais, pour l'instant, nous pouvons retenir que la structure optique est corrélée nomologiquement aux propriétés occurrentes.

Donc, en bref, la structure optique est corrélée nomologiquement aux propriétés occurrentes et les propriétés occurrentes sont corrélées nomologiquement à l'affordance. Mais la relation de corrélation est une relation transitive. Donc, par transitivité, la structure optique est corrélée nomologiquement à l'affordance. Or, nous avons dit ci-dessus que, si un phénomène physique était corrélé nomologiquement à un état de choses, ce phénomène physique transmettait l'information sur cet état de choses. Donc, la structure optique transmet l'information à propos de l'affordance.

Aussi, comme la perception d'une affordance consiste en la saisie de l'information correspondant à cette affordance et comme nous venons de voir que toute l'information requise est dans la structure optique, nous pouvons conclure que la perception d'une affordance consiste en la saisie de la structure optique. C'est au chapitre 5 que je vais montrer comment le système perceptuel saisit cette structure optique. Mais, pour mener ce dessein à terme, il faut d'abord déterminer ce qui caractérise la structure optique qui spécifie une affordance.

3.2.2. En quoi consiste la structure optique qui spécifie une affordance?

La structure optique spécifiant une affordance se présente comme une *combinaison invariante d'invariants* (Gibson, 1986, p. 140). Pour comprendre ce qu'est un invariant d'invariant et pour comprendre pourquoi la structure optique spécifiant une affordance se présente comme un invariant d'invariants, il faut d'abord se donner la peine de définir ce qui caractérise la structure optique qui spécifie une propriété occurrente, disons la forme d'un objet.

La structure optique spécifiant la forme d'un objet se présente non pas comme une combinaison invariante d'invariants (un invariant d'invariants), mais comme un simple invariant. Clarifions maintenant la notion d'invariant.

La lumière ambiante est réfléchie par les objets qui meublent l'environnement et cette lumière réfléchie est projetée vers un point d'observation. Lorsque la lumière est projetée

vers le point d'observation, elle se présente comme une *structure de perspectives*. En effet, l'environnement réfléchit la lumière avec une intensité variable et une fréquence variable, et c'est cette variation d'intensité et de fréquence qui *structure* la lumière projetée vers le point d'observation. La réflexion de la lumière par l'environnement varie d'intensité parce que les objets de l'environnement ne sont pas tous orientés de la même façon et parce qu'ils ont des capacités d'absorption variables. Et la variation des fréquences provient de surfaces ayant un degré de réflectance différent.

Lorsqu'elle est projetée sur l'œil de l'observateur, la structure de perspectives *isolée* ne transmet pas encore l'information à propos de la propriété. Plus précisément, la structure de perspectives ne renvoie pas de manière univoque à la propriété (elle n'est pas corrélée nomologiquement à la propriété); elle donne donc une information équivoque ou ambiguë sur la propriété. Pour que la structure optique renvoie de manière univoque à la propriété et qu'elle porte l'information à propos de cette propriété, on doit pouvoir dégager un *invariant* optique et cet invariant est dégagé sur la base d'une succession de structures de perspectives. Les structures de perspectives qui se succèdent à la faveur des mouvements de l'observateur sont différentes entre elles, mais elles ont aussi des caractéristiques communes ou identiques. C'est ce « quelque chose » de commun au sein des structures de perspectives qui correspond à l'invariant. L'invariant optique, comme on vient de le suggérer, porte l'information sur la propriété ou, autrement dit, il y renvoie de manière univoque (plus précisément : il est corrélé nomologiquement à la propriété). C'est en saisissant ou extrayant cet invariant que l'observateur perçoit la propriété. Afin de bien comprendre à quoi correspond l'invariant, nous allons nous appuyer sur un exemple.

Considérons la forme d'un objet et supposons que cette forme est carrée. Si l'observateur se déplace autour de cette surface, il va faire l'expérience d'une multitude de structures de perspectives successives. Ces structures de perspectives diffèrent les unes des autres, mais elles partagent néanmoins certaines caractéristiques. Par exemple, toutes ces structures de perspectives ont des angles intérieurs dont la somme égale 360 degrés. Nous avons donc ici une caractéristique invariante. Les structures de perspectives projetées par le carré ont certainement d'autres propriétés en commun. L'ensemble de ces caractéristiques communes correspond à ce que Gibson nomme « invariant ». C'est en saisissant cet invariant que

l'observateur perçoit une forme carrée, car, comme on l'a dit, cet invariant renvoie de manière univoque (est corrélé nomologiquement) à la surface carrée.

Cette description d'un invariant qui spécifie une propriété occurrente devrait nous permettre de comprendre comment Gibson caractérise la structure optique correspondant aux affordances, c'est-à-dire comme un invariant d'invariants.

La base occurrente de l'affordance, c'est-à-dire l'ensemble de propriétés occurrentes qui est à la base de l'affordance, se présente comme une combinaison *invariante* de propriétés occurrentes. Ainsi, pour qu'une chaise présente l'affordance « assieds-toi », les cinq propriétés occurrentes susmentionnées doivent être invariablement réunies ou combinées. Or, la structure optique qui porte l'information sur une propriété occurrente a été définie comme un invariant. Donc, la structure optique qui porte l'information sur une affordance se caractérise comme une combinaison invariante d'invariants ou, autrement dit, comme un invariant d'invariants.

3.2.3. La relation à l'effectivité dans la perception d'une affordance

Les affordances, comme on l'a dit, sont des propriétés relationnelles; elles sont en relation avec l'agent (ses effectivités). Comment ce rapport ou cette relation à l'agent se traduit-il dans la perception d'une affordance?

Lors de la perception, les propriétés occurrentes doivent être mesurées. C'est lors de la mesure de ces propriétés qu'une relation à l'effectivité est intégrée au sein de la perception. Plus précisément, la relation à l'effectivité est intégrée au sein de la perception lorsque les propriétés occurrentes sont mesurées relativement à une échelle de mesure qui est le corps ou les capacités physiques de l'agent.

Donc, la mesure des propriétés occurrentes de l'affordance peut être réalisée de deux façons : ou bien en rapportant les propriétés occurrentes aux dimensions du corps de l'animal, ou bien en les rapportant aux capacités physiques de l'animal. Le choix du cadre de référence est fonction du type de l'affordance. Les affordances du premier type sont qualifiées de « body-scaled affordances » et celles du second type sont qualifiées « d'action-

scaled affordances » (Fajen *et al.*, 2009)[37]. Il faudrait expliciter ces deux façons de rapporter l'affordance (en fait : ses propriétés occurrentes) à l'agent. Commençons par examiner l'affordance dont les propriétés occurrentes sont rapportées aux dimensions du corps de l'agent (body-scaled affordance).

Considérons une chaise. Parmi les propriétés constitutives de la possibilité de s'asseoir, il y a la hauteur de la chaise. En effet, pour qu'une chaise « offre » l'action de s'asseoir, il faut qu'elle ait une certaine hauteur. La mesure de cette hauteur, comme on l'a dit, n'est pas extrinsèque (par ex., en cm), mais intrinsèque : cette hauteur est mesurée à l'échelle des dimensions du corps de l'agent. Ainsi, l'unité de mesure pourrait être la hauteur des yeux de l'agent, qui est spécifiée optiquement, et la hauteur de la chaise serait rapportée à cette unité de mesure : elle (la hauteur de la chaise) serait mesurée comme étant une fraction de la hauteur des yeux (Fajen *et al.*, 2009). Autrement dit, lorsqu'elle est mesurée de manière intrinsèque, la mesure de la hauteur de la chaise se ramène au *ratio* de la hauteur de la chaise sur la grandeur de l'observateur.

Il est plus difficile de mesurer la propriété occurrente lorsqu'il s'agit de la mettre à l'échelle des capacités physiques. Autrement dit, les « action-scaled affordances » posent davantage de problèmes.

Ici, de manière analogue aux « body-scaled affordances », la mesure de la propriété occurrente est réalisée en établissant le ratio de la propriété occurrente sur la capacité physique. Le problème, toutefois, c'est que la propriété occurrente et les capacités physiques de l'agent sont apparemment « incommensurables ». Notre problème est donc celui-ci : trouver un point commun entre les propriétés occurrentes et les capacités physiques de l'agent de manière à pouvoir mettre les propriétés occurrentes à l'échelle des capacités physiques de l'agent. Ce problème est résolu à partir du moment où l'on réalise que la structure optique est beaucoup plus riche que l'on croit. On y trouve effectivement une information qui est parfaitement commensurable avec les capacités physiques de l'agent. Prenons la situation qui « afforde » le freinage. La capacité de freiner de l'agent est

---

[37] Il est fréquent que l'affordance soit en relation à la fois avec le corps de l'animal et avec ses capacités physiques. Dans ces circonstances, la mesure de l'affordance peut utiliser un cadre de référence mixte. Certaines des propriétés occurrentes seront mesurées par rapport aux corps de l'animal et les autres propriétés occurrentes seront mesurées par rapport à ses capacités physiques.

identifiable à une capacité à décélérer. Donc, pour mettre la situation environnementale à l'échelle de l'agent, il faudrait que l'information se rapportant à cette situation puisse se décliner en termes de décélération. Or, dans le flux optique donnant de l'information sur la situation environnementale, on retrouve précisément une information qui concerne la décélération. Plus précisément, il s'agit d'une information qui spécifie la décélération idéale, c'est-à-dire la décélération constante et non perturbée qu'un mobile doit effectuer pour arrêter à l'endroit voulu[38]. On peut donc mettre la décélération idéale, spécifiée dans la structure optique, à l'échelle de la capacité de décélérer de l'agent. Plus précisément, Fajen (2005) a montré que la décélération idéale était à l'échelle de la décélération maximale de l'agent. Ainsi, si la décélération idéale est perçue comme étant plus grande que la décélération maximale, la situation n'est pas perçue comme « affordant » l'action de freiner, tandis que si le ratio décélération idéale/décélération maximale est inférieur (ou égal) à 1/1, cette situation « afforde » l'action de freiner. C'est ainsi que la situation environnementale est mise à l'échelle des capacités physiques de l'agent.

Ce qui précède nous permet de voir que Stoffregen (2003) avait tort de dire que la définition des affordances de Turvey (1992) implique une théorie *indirecte* de la perception des affordances. Comme les propriétés occurrentes sont mises à l'échelle des dimensions du corps de l'agent ainsi qu'avec ses capacités physiques, ces propriétés occurrentes sont mises en relation avec l'effectivité de l'agent. Ici, il est possible de s'objecter et de faire valoir que les propriétés occurrentes de l'affordance ne sont pas mises en relation avec l'effectivité, mais avec les propriétés occurrentes de l'effectivité (dimensions du corps et capacités physiques). À cela, nous répondons ceci. Ces propriétés (le corps et les capacités physiques de l'agent), compte tenu qu'elles sont les propriétés occurrentes de l'effectivité (une disposition), sont corrélées nomologiquement à l'effectivité. Or, si elles sont corrélées nomologiquement à l'effectivité, ces propriétés possèdent l'effectivité. Donc, lorsque les propriétés occurrentes de l'affordance sont mises en relation avec les propriétés occurrentes de l'effectivité, elles sont aussi mises en relation avec cette effectivité.

---

[38] La décélération idéale spécifiée dans le flux optique correspond à une combinaison du taux du flux optique global (GOFR) et de τ (c-à-d, GOFR/2τ). τ spécifie le temps de l'impact (time-to-contact) avec l'objet et il correspond au ratio de l'angle optique de l'objet sur le taux de l'expansion optique de cet angle (Fajen, 2005).

Nous pouvons donc rajouter ce qui suit : la théorie des affordances de Turvey est compatible avec une théorie de la perception *directe* des affordances. À ce titre, l'objection de Stoffregen contre la théorie des affordances de Turvey n'est pas valable. Et, comme nous avons déjà donné une raison pour adopter la théorie de Turvey, à savoir qu'elle était la seule à respecter le fait que les affordances se présentent comme des potentialités ou des possibilités, nous pouvons dire, avec Turvey, que les affordances sont des propriétés de l'environnement et que ces propriétés sont des dispositions.

3.3. Résumé

Dans la première partie du chapitre 3, j'ai tenté de définir les affordances. Deux types de définitions s'opposent, et j'ai tenté de choisir mon camp. Pour les uns, les affordances sont des relations entre les propriétés occurrentes de l'objet et les propriétés occurrentes de l'agent. Pour les autres, les affordances sont des propriétés de l'environnement. J'ai présenté un représentant majeur de chacune de ces deux façons de voir les affordances. Turvey (1992) considère que les affordances comme des propriétés de l'environnement; Stoffregen (2003), comme des relations.

Pour Turvey, l'affordance est une propriété dispositionnelle de l'objet. Les dispositions sont des potentialités et, à ce titre, elles sont inobservables. Les affordances sont donc inobservables. Elles sont possédées par un objet lorsque cet objet est formé d'une combinaison invariante de propriétés occurrentes; l'affordance résulte de cette combinaison invariante de propriétés occurrentes. Ainsi, la chaise acquiert son affordance lorsque celle-ci contient les cinq propriétés occurrentes susmentionnées. Cependant, en tant que disposition, l'affordance n'est pas seulement redevable de cette combinaison invariante de propriétés. Elle est aussi redevable de l'effectivité. En effet, si la disposition à s'asseoir (l'effectivité) n'était pas une chose possible, la chaise « n'afforderait » pas l'action de s'asseoir. Donc, l'affordance entretient une relation essentielle à son effectivité (1ère caractéristique). Aussi, si elle est sous le coup d'une intention appropriée, l'affordance ne manque jamais de s'actualiser lorsqu'elle est jointe à son effectivité ou complément (3ème caractéristique). Pour finir, il faut dire que l'affordance existe même lorsqu'elle n'est pas jointe à son effectivité et que, donc, elle n'est pas actualisée (2ème caractéristique).

Après avoir caractérisé les affordances comme des dispositions, Turvey en donne une définition formelle. Je renvoie le lecteur au texte pour le détail de cette définition (p. 61).

Stoffregen (2003) définit les affordances comme des relations. Plus précisément, il considère que l'affordance est une propriété qui émerge de la relation entre les propriétés occurrentes de l'objet et les propriétés occurrentes de l'agent. Nous avons vu que cette façon de caractériser les affordances ne faisait pas l'affaire. Les affordances sont des possibilités d'action et, à ce titre, elles sont des potentialités. Or, la propriété qui émerge de la relation entre les propriétés occurrentes de l'objet et les propriétés occurrentes de l'agent n'est pas une potentialité, mais une propriété actuelle.

Ensuite, nous avons répondu aux objections que Stoffregen a adressées à Turvey. Ces objections sont au nombre de deux.

Premièrement, Stoffregen considère que la définition de Turvey est incomplète. Selon Stoffregen, la fonction de juxtaposition, qui est nécessaire pour comprendre que l'agent ne cherche pas à actualiser la multitude d'affordances qui se présentent à lui, n'est pas définie par Turvey. Donc, comme la définition de Turvey réfère à cette fonction de juxtaposition, la définition de Turvey est incomplète. Nous avons répondu à cette objection en faisant valoir que le filtre (la fonction de juxtaposition) pouvait être défini comme une intention.

La seconde objection de Stoffregen concerne la perception. Selon lui, en vertu de sa définition, Turvey serait contraint de comprendre la perception des affordances comme une perception indirecte, c'est-à-dire, comme une perception qui ne repose pas exclusivement sur la structure optique. Essentiellement, l'argument de Stoffregen est celui-ci : l'affordance implique une référence à l'effectivité; or, lorsque Turvey décrit la spécification de l'affordance par la structure optique, la relation à l'effectivité n'est aucunement spécifiée par la structure optique; donc, la relation à l'effectivité doit être ajoutée indirectement (mentalement) à la perception de l'affordance. Nous avons répondu à cette objection en faisant valoir que, lors de la spécification optique des propriétés occurrentes de l'affordance, ces dernières étaient mises à l'échelle (scaled) de l'effectivité. Cela fait en sorte que la spécification optique de l'affordance fait référence à la relation à l'effectivité.

Comme les deux objections de Stoffregen ont été réfutées et comme sa définition a été critiquée, nous avons convenu d'adopter la définition de Turvey : Les affordances sont des propriétés de l'environnement, plus précisément, des dispositions de l'environnement.

Dans la seconde partie du chapitre 3, nous nous sommes penchés sur la perception des affordances.

Pour une théorie directe de la perception, percevoir une propriété, c'est saisir la structure optique qui porte l'information sur la propriété en question. Donc, percevoir une affordance, c'est saisir la structure optique qui porte l'information sur l'affordance. Deux questions se posent ici : 1) La structure optique porte-t-elle l'information sur l'affordance? 2) Comment le système perceptuel peut-il saisir la structure optique qui porte l'information sur l'affordance? Je réponds à la seconde question dans le chapitre 5 et j'ai tenté de répondre à la première question au cours du présent chapitre. Pour y répondre, j'ai d'abord énoncé les conditions qu'un signal (optique ou autre) doit satisfaire pour pouvoir transmettre l'information à propos d'une propriété. Je me suis appuyé sur Dretske (1981/1999) pour trouver ces conditions. Ensuite, j'ai tenté de savoir si le signal (la structure optique) correspondant aux affordances satisfaisait ces conditions.

La première condition est que la quantité d'information associée au signal doit être aussi grande que la quantité d'information associée à l'état de choses pour lequel nous voulons de l'information. La deuxième condition est que cet état de choses, et non pas un autre, doit être à l'origine du signal. La troisième condition est que la quantité d'information associée au signal doit être la quantité d'information associée à l'état de choses. Après avoir trouvé ces conditions, nous avons trouvé une définition de la transmission d'information par un signal qui satisfaisait ces conditions : Un signal transmet l'information portant sur un état de choses si et seulement si le signal est corrélé nomologiquement à l'état de choses[39]. Donc, pour savoir si la structure optique transmet l'information à propos d'une affordance, il suffit de s'assurer que la structure optique est corrélée nomologiquement à l'affordance. C'est ce que nous avons tenté de faire en montrant, d'une part, que la combinaison des propriétés occurrentes était corrélée nomologiquement à l'affordance et, d'autre part, en montrant que la structure optique était corrélée nomologiquement à cette combinaison de propriétés

[39] La corrélation doit être de 1.

occurrentes. Donc, par transitivité, la structure optique est corrélée nomologiquement à l'affordance; elle transmet l'information sur l'affordance.

Après avoir montré que la structure optique transmettait l'information sur l'affordance, j'ai cherché à savoir ce qui caractérisait cette structure optique. Celle-ci se caractérise comme un invariant d'invariants. Cette structure optique spécifie une combinaison invariante de propriétés occurrentes. Donc, cette structure optique est une combinaison invariante des structures optiques qui spécifient chacune des propriétés occurrentes. Nous avons vu que chacune des structures optiques (celles qui spécifient une propriété occurrente) se caractérise comme un invariant optique. Donc, la structure optique qui spécifie la combinaison invariante de propriétés occurrentes et, du fait même, l'affordance, se caractérisent bel et bien comme une combinaison invariante d'invariants ou, pour faire plus court, comme un invariant d'invariants.

Dans la dernière section du chapitre, j'ai donné deux exemples où l'on peut constater que les propriétés occurrentes sont mises à l'échelle de l'effectivité. Le premier exemple mettait une propriété occurrente à l'échelle des dimensions du corps de l'agent (body-scaled affordance), et le deuxième exemple mettait une propriété occurrente à l'échelle des capacités physiques de l'agent (action-scaled affordance).

CHAPITRE IV

UNE DÉFENSE DE LA PERCEPTION DIRECTE

Comme je l'ai annoncé au chapitre 3, le chapitre 4 tente de réfuter l'une des critiques majeures qui a été adressée à la position de Gibson, la critique que Fodor et Pylyshyn présentèrent dans leur article « How direct is visual perception? Some reflections on Gibson's "ecological approach". » (1981) La réfutation de cette critique sera aussi pour moi une occasion d'approfondir l'interprétation que j'ai donnée plus tôt de l'approche de Gibson.

Fodor et Pylyshyn présentent deux objections contre la possibilité d'une perception directe, c'est-à-dire non inférentielle. L'objectif du présent chapitre est de présenter ces objections et de les réfuter. Cependant, avant d'examiner ces deux objections, j'aimerais attirer l'attention sur l'une des thèses que Fodor et Pylyshyn défendent dans leur article. Ces derniers soutiennent que Gibson ne parvient pas à définir le domaine de ce qui est directement perceptible. Je compte réfuter leurs arguments et montrer que Gibson parvient à circonscrire le domaine des propriétés directement perceptibles. Ce chapitre comprend donc trois parties. Premièrement, il s'agit de montrer que Gibson est parvenu à circonscrire le domaine des propriétés directement perceptible. Deuxièmement, il s'agit de montrer que la première objection de Fodor et Pylyshyn contre la perception directe ne tient pas la route. Troisièmement, il s'agit de montrer que leur deuxième objection contre cette même perception directe ne tient pas davantage la route.

4.1 Le domaine des propriétés directement perceptibles

Fodor et Pylyshyn trouvent chez Gibson quatre critères permettant de distinguer ce qui est directement perceptible. Ils critiquent ces quatre critères. Je suis d'accord avec leurs critiques

en ce qui concerne deux de ces quatre critères, mais je crois que les deux autres critères m'apparaissent valables pour définir ce qui est directement perceptible.

Le critère phénoménologique : les propriétés directement perçues sont des propriétés ayant un caractère phénoménologique[40]. Fodor et Pylyshyn développent trois arguments contre l'identification des propriétés directement perçues aux propriétés phénoménologiques.

Premièrement, Fodor et Pylyshyn remarquent que plusieurs propriétés tenues par Gibson pour être directement perceptibles ne sont pas données avec un caractère phénoménologique. Comme exemple de propriété directement perçue mais inaccessible phénoménologiquement (inaccessible à la conscience), Fodor et Pylyshyn énumèrent l'inclinaison des surfaces, le flux de la texture (lorsqu'il y a mouvement), le bord texturé occlusif en mouvement (moving occluding texture edge), etc. Donc, comme ces propriétés sont tenues par Gibson pour être directement perçues et comme elles ne sont pas données phénoménologiquement, le critère phénoménologique pour définir ce qui est directement perceptible ne tient pas la route. Ce premier argument m'apparaît convaincant.

Deuxièmement, pour montrer que les propriétés directement perceptibles ne sont pas des propriétés phénoménologiques, Fodor et Pylyshyn montrent que l'idée même de propriétés phénoménologiques est en soi une idée douteuse. Pour montrer qu'il n'y a pas de propriétés phénoménologiques, il leur suffit de montrer que le donné phénoménologique lui-même dépend ou présuppose une quantité de données non phénoménologiques. À cet égard, Fodor et Pylyshyn réfèrent à une expérience de Hess (1975), où ce dernier montre que les propriétés phénoménologiques d'un visage dépendent d'aspects du visage qui passent inaperçus par l'observateur. Par exemple, le caractère attirant d'un visage, un trait phénoménologique, dépendrait du diamètre des pupilles, un trait presque inaccessible phénoménologiquement. Cet argument aussi m'apparaît convaincant.

Troisièmement, Fodor et Pylyshyn font valoir que, pour qu'il soit acceptable d'identifier les propriétés directement perceptibles aux propriétés phénoménologiques (ou propriétés données à la conscience), il faudrait à tout le moins entrevoir la possibilité d'un mécanisme

---

[40] Il existe plusieurs types de critères permettant de distinguer un objet ou une propriété. Ce critère-ci, le critère phénoménologique, est définitionnel; il *identifie* les propriétés directement perceptibles à ce qui est phénoménologique.

direct pouvant être à la base de la perception des propriétés phénoménologiques. Or, dans le cas de propriétés phénoménologiques comme les affordances, Fodor et Pylyshyn ne voient pas comment un tel mécanisme pourrait être développé. Cet argument ne m'apparaît pas valable. Comme on l'a vu au chapitre précédent, l'affordance est perçue à travers la saisie d'un invariant d'invariants et je ne vois pas pourquoi un mécanisme direct pouvant saisir ou extraire ce type d'invariant optique est en soi quelque chose d'inaccessible. En fait, plus tard dans ma thèse (chapitre 5), je vais tracer les grandes lignes de ce genre de mécanisme. Ce mécanisme sera inspiré du réseau de neurones récurrent simple de Elman.

Le critère comportemental : L'autre critère que Fodor et Pylyshyn tirent de Gibson est celui-ci : ce qui est directement perçu correspond à ce à quoi le système perceptuel répond. Évidemment, pour déterminer ce à quoi répond le système perceptuel, il faut d'abord définir ce qu'on entend par « système perceptuel ». Ainsi, si on le définit comme étant la rétine, on pourra observer que ce système perceptuel répond à la lumière. La question qu'il faut se poser est donc celle-ci : à quoi correspond le système perceptuel chez Gibson. Bien que Gibson donne quelques indications d'ordre anatomique pour définir le système perceptuel, pour l'essentiel, il le définit de manière fonctionnelle. Selon Fodor et Pylyshyn, Gibson définit fonctionnellement le système perceptuel en se contentant de dire qu'il (le système perceptuel) répond à ce qui est directement perceptible. Or, ce qui est directement perceptible est précisément la notion que l'on tente de définir en premier lieu. Le système perceptuel reste donc non défini chez Gibson. Or, comme on l'a dit, pour définir les propriétés directement perceptibles à partir de ce à quoi le système perceptuel répond, il faut se doter d'une définition du système perceptuel. Le critère comportemental n'est donc pas valable.

Le critère écologique : En lisant Gibson, on constate qu'il soutient que les propriétés directement perceptibles sont des propriétés écologiques. Le problème, selon Fodor et Pylyshyn, c'est que Gibson ne parvient pas à définir adéquatement le terme « écologique ». Plus précisément, selon Fodor et Pylyshyn, Gibson définirait les propriétés écologiques comme des propriétés perceptibles. Or, Gibson ne peut pas définir de la sorte les propriétés écologiques, si ces dernières sont supposées définir ce qui est directement perceptible. Comme avec le second critère, Gibson nous entraînerait ici dans un cercle vicieux.

Cet argument de Fodor et Pylyshyn est valable seulement si Gibson définit effectivement les propriétés écologiques comme des propriétés directement perceptibles. Or, selon Turvey *et al.* (1981), cela n'est pas juste; Gibson définit les propriétés écologiques comme des propriétés ayant une signification pour les actions de l'animal. Autrement dit, les propriétés écologiques sont des propriétés qui se situent à une certaine échelle de grandeur, à une échelle de grandeur significative pour les actions de l'animal. Donc, définir les propriétés directement perceptibles comme des propriétés écologiques n'implique aucun cercle vicieux. Je considère donc que ce critère — les propriétés directement perceptibles sont des propriétés écologiques ou, autrement dit, des propriétés qui se situent à une échelle de grandeur significative pour les actions de l'animal — est un critère valable pour identifier ce qui est directement perceptible.

Cependant, le critère écologique est nécessaire, mais non suffisant, pour identifier les propriétés directement perceptibles. Il est nécessaire car il n'y a pas de propriétés directement perceptibles qui ne soient une propriété écologique, c'est-à-dire une propriété située à une échelle de grandeur significative pour les actions de l'animal. Mais ce critère n'est pas suffisant, car il y a des propriétés qui sont situées à une échelle de grandeur significative pour les actions de l'animal et ces propriétés ne sont pas directement perceptibles. Pensons par exemple à la propriété de l'interrupteur de fermer (ou ouvrir) la lumière. Cette propriété n'est pas directement perceptible et elle est pourtant située à une échelle de grandeur qui est significative pour les actions de l'agent. Il faut donc ajouter un autre critère au critère écologique. Cet autre critère, qui est le quatrième sur la liste des critères que Fodor et Pylyshyn critiquent, est nécessaire et suffisant. Mais comment puis-je avoir deux critères tandis que l'un est nécessaire et l'autre, nécessaire et suffisant. Cela est possible car ces deux critères ne sont pas complètement distincts : le second critère, celui qui est nécessaire et suffisant, comprend en lui le premier critère, celui qui est seulement nécessaire et qui pose que les propriétés directement perceptibles sont des propriétés écologiques (situées à une échelle de grandeur significative pour les actions de l'animal).

Le critère écologique-nomologique : Fodor et Pylyshyn trouvent chez Gibson ce quatrième critère : Les propriétés directement perceptibles sont celles qui entrent dans les lois de l'optique écologique. Les lois de l'optique écologique connectent nomologiquement une

propriété écologique (une propriété du monde qui se situe à une échelle de grandeur significative pour les actions de l'animal) à la structure lumineuse que la propriété écologique reflète et projette vers un point d'observation. Donc, autrement dit, le critère écologique-nomologique nous dit ceci : les propriétés directement perceptibles sont des propriétés écologiques qui réfléchissent et structurent nomologiquement la lumière ambiante. Une précision doit cependant être faite ici : lorsqu'une propriété écologique réfléchit et structure la lumière d'une certaine façon, cette propriété écologique est la seule à pouvoir structurer la lumière de cette façon. Autrement dit, une structuration donnée de la lumière ne peut pas avoir plus d'une cause (une propriété écologique). On doit pouvoir dire : si une structure optique X, alors une propriété écologique Y. Cette précision apportée au quatrième critère est importante si l'on veut pouvoir rendre compte du fait que la structure optique porte l'information sur la propriété écologique. En effet, comme on l'a vu au chapitre précédent, pour que X porte l'information sur Y, on doit pouvoir dire : Si X, alors Y.

Ce quatrième critère, contrairement au précédent, est nécessaire et suffisant pour circonscrire le domaine des propriétés directement perceptibles. Il est nécessaire, car si une propriété est directement perceptible, cette propriété est une propriété écologique qui réfléchit nomologiquement la lumière. Et il est suffisant, car si une propriété écologique réfléchit nomologiquement la lumière, elle est directement perçue.

Comme je l'ai dit plus haut, ce dernier critère a été critiqué par Fodor et Pylyshyn.

Fodor et Pylyshyn remarquent d'abord que cette façon de définir les propriétés directement perceptibles limite significativement le nombre des propriétés directement perceptibles. Par exemple, ils font remarquer que la propriété d'une toile d'être peinte par De Vinci ne saurait être perçue directement puisqu'il n'existe pas de lois qui connectent cette propriété à la structure lumineuse. Fodor et Pylyshyn ont raison; il n'existe pas de lois qui connectent la structure lumineuse projetée par cette toile au fait que cette toile a été peinte par un personnage historique dont le nom était « Leonard De Vinci ». Il est donc exclu que nous puissions percevoir directement cette propriété. Cependant, il est raisonnable de croire que je puisse percevoir directement une propriété qui s'apparente de très près à la propriété d'être peinte par De Vinci; je peux percevoir que cette toile a été peinte par un individu dont les capacités et le tempérament correspondent aux capacités et au tempérament de De Vinci. La

structure lumineuse réfléchie par les traits de peinture est effectivement connectée à la propriété d'être peinte par un individu dont la constitution psychomotrice est identique à la constitution psychomotrice de De Vinci. Je m'explique.

À quoi se ramène la propriété d'une toile d'être peinte par un individu ayant les capacités psychomotrices X? Elle se ramène évidemment au fait qu'un individu ayant cette constitution psychomotrice a peint la toile. Donc, lorsque l'on se demande si la structure lumineuse spécifie (est corrélée nomologiquement à) la propriété d'une toile d'être peinte par un individu ayant une capacité psychomotrice X, on se demande en fait si la structure lumineuse est connectée nomologiquement au fait qu'un individu ayant cette capacité psychomotrice a peint cette toile. La structure lumineuse pourrait renvoyer à ce fait si l'on pouvait montrer que l'individu ayant cette capacité psychomotrice a pour ainsi dire laissé de lui-même dans les traits de peinture qu'il a déposés sur la toile. En effet, il est clair que la structure optique est corrélée nomologiquement aux traits de peinture. Or, l'individu ayant cette capacité psychomotrice a précisément laissé de lui-même dans les traits de peinture qu'il a déposés sur ses toiles. Ces traits de peinture sont en effet corrélés nomologiquement aux particularités psychomotrices (capacités de peindre et tempérament) de l'individu.

Donc, le critère qui vient d'être donné à propos des propriétés directement perceptibles — les propriétés directement perceptibles sont des propriétés écologiques qui réfléchissent nomologiquement la lumière — ne limite pas outre mesure le nombre des propriétés directement perceptibles.

Nous avons donc en notre possession un critère nécessaire et suffisant pour définir ce qui est directement perceptible : les propriétés directement perceptibles sont les propriétés écologiques qui réfléchissent ou structurent nomologiquement la lumière ambiante. Comme je l'ai dit, Fodor et Pylyshyn s'opposent au critère écologique-nomologique. À la lumière de ce qui a été dit jusqu'à présent, je vais montrer que leur argument ne tient pas la route. Mais examinons d'abord leur argument.

Ce dernier critère — les propriétés directement perceptibles sont des propriétés écologiques qui réfléchissent et structurent nomologiquement la lumière — dépend de l'existence de lois ressortant à l'optique écologique. En effet, la réflexion et la structuration de la lumière par la propriété écologique sont nomologiques parce qu'elles peuvent faire

l'objet de lois de l'optique écologique. Donc, pour déterminer les propriétés qui réfléchissent et structurent nomologiquement la lumière, il faut pouvoir identifier les lois de l'optique écologique. Or, selon Fodor et Pylyshyn, Gibson ne donne aucun critère permettant d'identifier les lois de l'optique écologique. En fait, Gibson définirait les lois de l'optique écologique en disant qu'elles se rapportent aux propriétés écologiques. Or, selon Fodor et Pylyshyn, cette référence aux propriétés écologiques ne fait pas l'affaire, puisque, comme on l'a vu plus tôt, Gibson définirait les propriétés écologiques comme des propriétés directement perceptibles, ce qui est la notion problématique du départ.

À la lumière de ce qui a été dit plus tôt, il n'est pas difficile de voir ce qui ne va pas avec l'argument de Fodor et Pylyshyn : Gibson ne définit pas les propriétés écologiques comme des propriétés directement perceptibles, mais comme des propriétés qui se situent à une échelle de grandeur significative pour les actions de l'animal. Les lois de l'optique écologique ne se définissent donc pas en faisant référence au concept qui est problématique en premier lieu, c'est-à-dire au concept « directement perceptible ». Les lois de l'optique écologique sont donc clairement définies; notre dernier critère est valable : les propriétés directement perceptibles sont des propriétés écologiques qui réfléchissent et structurent nomologiquement la lumière.

Donc, il semblerait que, contrairement à ce que croient Fodor et Pylyshyn, le domaine des propriétés directement perceptibles soit clairement défini par Gibson. Mais le cœur du texte de Fodor et Pylyshyn ne porte pas sur la possibilité de circonscrire le domaine des propriétés directement perceptibles, mais sur la possibilité de la perception directe. Fodor et Pylyshyn critiquent cette dernière possibilité. Les deux prochaines sections décrivent et tentent de réfuter ces critiques.

4.2. La 1ère objection de Fodor et Pylyshyn

La théorie de la perception directe de Gibson est somme toute assez simple et elle a été en grande partie décrite dans le chapitre précédent. Elle se base sur la théorie de l'information. Grosso modo, voici comment on peut résumer l'argument qui justifie la perception directe.

Si on considère que percevoir une propriété écologique c'est acquérir de l'information à propos de la propriété écologique via un système perceptuel, la perception d'une propriété

peut consister à détecter (via un système perceptuel) un signal qui porte ou contient en soi de l'information sur cette propriété. Or, comme on l'a vu plus tôt, un signal porte de l'information sur une propriété lorsque le signal est corrélé nomologiquement à la propriété. Donc, pour percevoir une propriété, il suffit de détecter un phénomène physique (un signal) qui est corrélé nomologiquement à la propriété. Or, on vient de voir que la structure optique est corrélée nomologiquement à la propriété écologique. Donc, pour percevoir les propriétés écologiques, il suffit de détecter la structure optique. Or, la détection de la structure optique par le système perceptuel est *directe*, non médiatisée par des inférences[41]. Donc, la perception des propriétés écologiques est directe. Fodor et Pylyshyn n'acceptent pas l'entièreté de cet argument.

Ils acceptent l'idée que nous détectons immédiatement ou directement la structure optique. Plus précisément, comme ils identifient « détection directe » et « transduction », ils acceptent l'idée que la structure optique est transduite. Mais, ce qu'ils n'acceptent pas, c'est que la perception de la propriété écologique consiste en une détection directe de la structure optique. Ils n'acceptent pas cela, quand bien même la structure porterait l'information sur la propriété écologique[42]. Donnons un caractère plus précis à notre compte rendu de ce que Fodor et Pylyshyn n'acceptent pas. Supposons que S_1 (la structure optique) soit corrélé nomologiquement à S_2 (la propriété écologique) et que, donc, S_1 contient de l'information sur S_2. Percevoir S_2, c'est saisir le contenu « S_2 » ou, autrement dit, c'est être dans un état neuronal qui possède S_2 comme contenu. Donc, ce que Fodor et Pylyshyn n'acceptent pas, c'est que nous puissions saisir le contenu « S_2 » (être dans un état neuronal qui possède S_2 comme contenu) en détectant S_1. Pour comprendre leur refus à accepter cela, il faut revenir sur la théorie de l'information.

Pour Gibson, « S_1 contient de l'information sur S_2 » signifie qu'il y a de l'information portant sur S_2 *dans* S_1. L'état physique S_1 est donc en quelque sorte un contenant au sein duquel on retrouve un contenu, l'information portant sur S_2. Fodor et Pylyshyn qualifient cette analyse de l'information de réifiante. Lorsque l'on réifie ainsi l'information, la

---

[41] C'est dans le prochain chapitre que je vais montrer que la détection de la structure optique est un processus direct.

[42] Lors de la présentation de leur seconde objection, comme nous allons le voir, Fodor et Pylyshyn nient que la structure optique porte une information *complète* sur la propriété écologique.

perception de la propriété écologique (la saisie du contenu « S_2 ») se ramène effectivement à la détection de S_1, puisque S_1 contient littéralement de l'information sur S_2.

Le problème, soutiennent Fodor et Pylyshyn, c'est que la théorie de l'information ne permet pas de réifier ainsi l'information; elle ne permet pas d'en faire un contenu inclus dans un état de choses physique. Rappelons notre analyse de l'information.

S_1 contient de l'information sur S_2 si, et seulement si, S_1 est corrélé nomologiquement à S_2. Donc, « contient de l'information sur X » signifie seulement « est corrélé nomologiquement à X ». Autrement dit, notre analyse de l'information comprend « contient de l'information sur X » comme une propriété relationnelle. On ne peut donc pas, selon cette analyse de l'information, réifier l'information, en faire un contenu au sein d'un contenant. Ce que nous dit la théorie de l'information, ce n'est pas qu'il y a de l'information *dans* S_1, mais simplement que S_1 est corrélé à S_2. Donc, selon cette analyse de l'information, le transducteur, lorsqu'il détecte directement S_1, n'a pas S_2 comme contenu, puisque S_1 ne contient pas en lui l'information sur S_2. Donc, la perception de S_2 (la saisie du contenu « S_2 ») ne peut pas être identifiée à la détection directe de S_1.

Pourtant, nous percevons S_2, nous possédons un état neuronal qui a S_2 comme contenu. Comment pouvons-nous acquérir un tel état neuronal, si la détection de S_1 ne le fournit pas? En fait, selon Fodor et Pylyshyn, l'analyse de l'information qui vient d'être donnée — l'information comme propriété relationnelle — nous obligerait à opter pour une conception inférentielle de la perception. Avec l'analyse de l'information qui vient d'être donnée, S_1 (la structure informative ou indicative) peut mener à un état neuronal ayant S_2 comme contenu à condition que le système cognitif connaisse la corrélation entre S_1 et S_2. Plus précisément, si mon transducteur est dans l'état S_1 et si je connais la corrélation entre S_1 et S_2, je peux *inférer* que S_2. Je connais donc S_2 de manière indirecte ou inférentielle. La forme de l'inférence pourrait avoir la forme suivante[43] :

---

[43] La forme de cette inférence est arbitrairement choisie. En fait, selon Turvey *et al.* (1981), pour être plus près des théories cognitives sur la perception, l'inférence devrait être une inférence à la meilleure explication. La forme de l'inférence devrait donc être de type abductive. Mais la forme de l'inférence est sans implication sur les considérations qui vont suivre. Je supposerai donc, afin de simplifier mon propos, que l'inférence démonstrative décrite à l'instant est le type d'inférence qui est proposé par les inférentialistes en matière de perception.

S_1 (état du transducteur)

$S_1 \rightarrow S_2$ (connaissance de la corrélation entre S_1 et S_2)

S_2 (conclusion)

Pourquoi le modèle inférentiel fonctionne-t-il? Parce que, contrairement à l'état du transducteur, l'état neuronal qui réalise la connaissance de la corrélation entre S_1 et S_2 possède un contenu (ou un sens ou une signification), à savoir le contenu « $S_1 \rightarrow S_2$ ». En effet, l'état neuronal qui réalise la conclusion possède un contenu (S_2), et si l'état neuronal qui réalise la connaissance de la corrélation ne possédait pas le contenu susmentionné, il ne serait pas possible d'inférer cette conclusion (un état neuronal ayant S_2 comme contenu ou signification).

Donc, l'état neuronal qui réalise la connaissance de la corrélation n'est pas seulement dans une relation informationnelle avec la corrélation entre S_1 et S_2; contrairement à l'état du transducteur, il ne « signifie » pas simplement « est corrélé à ... »; il possède un contenu. Mais d'où provient le contenu (ou le sens ou la signification) de l'état neuronal qui réalise la connaissance de la corrélation entre S_1 et S_2? Dans leur texte de 1981, Fodor et Pylyshyn ne fournissent pas de réponse à cette question. Cependant, dans *Psychosemantics* (1987), Fodor donne une réponse à cette question, et cette réponse repose en grande partie sur la théorie de l'information. Dans ce texte, Fodor nous dit que l'état neuronal acquiert son contenu par une relation informationnelle avec un état de choses extérieur. Cependant, une simple relation informationnelle ne suffit pas pour donner un contenu à une structure indicative; il en faut davantage. Dretske (1988) soutient aussi que, pour qu'une structure indicative (ou informative) acquière un contenu (ou un sens ou une signification), il faut davantage qu'une relation informationnelle entre la structure indicative et la propriété. Je vais reprendre ici la théorie de Dretske.

Dretske soutient que, pour qu'une structure acquière le contenu « X est P », il faut, et il suffit, 1) que la structure soit dans une relation informationnelle ou d'indication avec l'état de choses « X est P » et 2) que la structure en question ait la *fonction* d'indiquer l'état de choses « X est P ». Donc, si, en plus d'être dans une relation informationnelle ou d'indication avec un état de choses, une structure a la fonction d'indiquer cet état de choses, cette structure

possède cet état de choses pour contenu. Donc, l'état neuronal qui réalise la connaissance de la corrélation entre S_1 et S_2 possède « $S_1 \rightarrow S_2$ » comme contenu 1) si cet état neuronal est dans une relation informationnelle ou d'indication (est corrélé nomologiquement) avec la corrélation entre S_1 et S_2 et 2) s'il possède la fonction d'indiquer cette corrélation.

Comme on l'a dit, l'état du transducteur est corrélé nomologiquement avec la propriété écologique S_2. Donc, pour pouvoir avoir « S_2 » comme contenu ou sens ou signification, il lui manque une seule chose : il faut qu'il ait la fonction d'indiquer la propriété « S_2 ». Or, l'évolution (Millikan, 1984) ou l'apprentissage (Dretske, 1988) peuvent donner à l'état du transducteur la fonction d'indiquer ce qu'il indique. L'état du transducteur ou détecteur peut donc avoir « S_2 » comme contenu ou signification. Lorsque le transducteur détecte S_1, la saisie du contenu « S_2 » (la perception de S_2) est réalisée.

Donc, à condition que le détecteur (ou transducteur) ait la fonction d'indiquer S_2, la perception de S_2 est effectivement identifiable à la détection de S_1. L'objection de Fodor et Pylyshyn est par le fait même réfutée. Cependant, il faut modifier sensiblement le modèle de Gibson. Pour qu'il y ait perception de S_2, il ne suffit pas qu'il y ait détection de S_1. Il faut que, en plus de pouvoir détecter S_1, le détecteur ait la fonction d'indiquer S_2.

Nous faisons donc face à deux modèles de la perception. Le premier est direct et l'autre est inférentiel. Mon choix s'arrête sur le modèle direct, inspiré de Gibson.

Tout d'abord, à la défense du modèle direct, il y a l'argument de la simplicité. L'état du transducteur et l'état neuronal issu de l'inférence sont tous deux des états qui ont « S_2 » comme contenu, la seule différence étant que l'état neuronal issu d'une inférence fait suite à un processus neuronal et computationnel compliqué, tandis que l'état du transducteur fait suite à la transduction, un processus infiniment plus simple. Donc, le modèle de la perception directe est plus simple que le modèle de la perception indirecte. Or, le fait qu'un modèle soit plus simple qu'un autre devrait nous inciter à privilégier le modèle plus simple, à condition, bien sûr, que toutes choses soient égales par ailleurs. Or, *prima facie*, le modèle direct est aussi explicatif que le modèle inférentiel (la clause *ceteris paribus* est respectée). Donc, le modèle direct devrait être privilégié.

Ce dernier argument n'est pas inattaquable. L'une de ses prémisses pourrait être fausse. Il s'agit de la prémisse qui pose qu'entre deux modèles antagonistes, on doive privilégier le

modèle le plus simple. En effet, il se pourrait que la nature se soit donnée de la peine pour fabriquer les mécanismes qui sont à la base de nos fonctions cognitives et donc, qu'entre deux modèles antagonistes, il faille privilégier le modèle le plus compliqué. À ce titre, il se pourrait que le modèle direct ne soit pas celui qui soit à la base de la perception.

Mais, quoi qu'il en soit, les propos tenus plus haut permettent à tout le moins d'affirmer que l'objection de Fodor et Pylyshyn contre le modèle direct n'est pas valide. L'objection de Fodor et Pylyshyn, je le rappelle, consiste à dire que l'état du transducteur (le signal S_1) ne peut pas avoir la propriété S_2 comme contenu, quand bien même le signal S_1 « contient » l'information sur S_2. Nous appuyant sur Dretske (1988), nous avons montré qu'à condition d'avoir la fonction d'indiquer S_2, S_1 possède le contenu S_2.

Mais Fodor et Pylyshyn présentent une autre objection au modèle direct de Gibson. Je me tourne maintenant vers cette deuxième objection.

4.3. La 2ème objection de Fodor et Pylyshyn : l'objection du « stimulus effectif »

Comme on l'a vu, Fodor et Pylyshyn estiment qu'il est nécessaire d'introduire un processus inférentiel pour passer de l'état du transducteur à un état neuronal ayant la propriété écologique comme contenu ou signification. Mais Fodor et Pylyshyn croient que l'usage d'inférence est aussi nécessaire à un autre moment du processus perceptuel.

Selon Fodor et Pylyshyn, les propriétés qui sont corrélées nomologiquement à la structure lumineuse sont en nombre limité, en nombre beaucoup plus limité que le nombre des propriétés du « layout »[44]. Donc, la quantité d'information contenue dans la structure optique est moins grande que la quantité d'information associée au « layout ». Or, la perception possède un contenu qui porte sur le « layout ». La perception possède donc davantage d'information que l'information comprise dans la structure lumineuse. Il faut donc que le contenu de la perception ait reçu cet excédent d'information de quelque part. La réponse de Fodor et Pylyshyn est de dire que la perception reçoit cet excédent d'information d'un processus inférentiel. Ainsi, si je perçois une tasse de café, le contenu de ma perception comprend une surface antérieure et une surface postérieure, et la structure lumineuse ne me

[44] Le « layout » correspond à l'ensemble des propriétés sur lesquelles porte la perception.

donne pas d'informations sur la surface postérieure. Comment puis-je avoir une perception dont le contenu excède ainsi le contenu informationnel de la structure lumineuse? La réponse de Fodor et Pylyshyn : ce sont des inférences qui viennent donner une surface postérieure au contenu de ma perception.

Il faut souligner que Fodor et Pylyshyn sont d'accord pour dire que si l'échantillonnage d'inputs lumineux n'est pas contraint, alors cet échantillonnage peut contenir autant d'information que la perception. Ainsi, si je me promène autour de la tasse de café, alors la quantité d'information de l'échantillonnage lumineux est comparable à la quantité d'information associée à la tasse (le « layout »). Donc, l'introduction d'inférences est justifiée, non pas en présence d'un stimulus illimité ou non contraint, mais seulement en présence d'un stimulus contraint. La question est alors de savoir si, lors d'une perception, le *stimulus effectif*, c'est-à-dire le stimulus auquel est effectivement soumis l'observateur, est contraint ou non. Autrement dit, lorsque l'observateur perçoit le « layout », se promene-t-il autour de l'objet de manière à recevoir un stimulus lumineux qui est aussi riche informationnellement que le « layout »? Ou, au contraire, le stimulus effectif possède-t-il moins d'information que l'information associée au « layout »? La thèse de Fodor et Pylyshyn, comme on vient de le dire, consiste à dire que la quantité d'information du stimulus effectif est moins grande que la quantité d'information associée au « layout ».

Essayons maintenant de préciser ce que Fodor et Pylyshyn veulent dire lorsqu'ils disent que la structure optique (le stimulus optique) contient moins d'information que l'information générée par l'occurrence du « layout ».

La théorie de l'information pose cette équation : information transmise par le signal (le stimulus effectif) = information générée à la source (pour nous : le « layout ») — équivoque. Donc, si le stimulus effectif (la structure lumineuse) contient moins d'information que l'information générée par l'occurrence du « layout », c'est que ce signal (le stimulus effectif) est équivoque. En disant qu'un signal est équivoque, nous disons que ce signal renvoie à plusieurs causes possibles; qu'il peut être causé par des événements différents. Un signal équivoque contient néanmoins un contenu informationnel (ce que le signal indique). Son contenu informationnel comprend la disjonction des événements qui peuvent le causer. Pour nous, cela signifie que le stimulus effectif ne porte pas l'information sur un « layout »

particulier mais sur une disjonction de « layouts ». Par exemple, si un stimulus effectif peut être causé par deux « layouts », le contenu informationnel de ce stimulus effectif sera: « layout n° 1 ou layout n° 2 ». Donc, lorsque Fodor et Pylyshyn nous disent que le stimulus effectif contient moins d'information que l'information générée par l'occurrence du « layout », ils disent que le signal (le stimulus effectif) indique (possède pour contenu informationnel) une disjonction de « layouts ». Soulignons que ceci est parfaitement intuitif. Si je sais que Sophie *ou* Marie viendra à la soirée, je sais *moins* de choses que si je sais que Marie viendra.

Pour éliminer cette équivoque et, donc, compléter l'information du signal, Fodor et Pylyshyn affirment qu'il faut procéder à des inférences. Gibson ne croit pas que l'on doive recourir à des inférences pour compléter l'information fournie par le stimulus effectif, car il croit que le stimulus effectif spécifie ou renvoie de manière univoque au « layout n° 1 » et donc que la structure lumineuse contient *toute* l'information sur le « layout n° 1 ». La thèse voulant que le stimulus effectif comprenne toute l'information sur le « layout » (ou y renvoie de manière univoque) est évidemment une thèse empirique. Autrement dit, on décide de sa vérité ou de sa fausseté en tâchant de voir si, pour chaque perception, le « layout » est spécifié de manière univoque par la structure optique.

Gibson a conduit plusieurs analyses de la structure lumineuse et il a montré au cours de ces analyses que plusieurs « layouts » étaient spécifiés par la structure lumineuse. Ainsi, il a montré qu'avec très peu de mouvements, c'est-à-dire sans se déplacer derrière l'objet, l'observateur pouvait être stimulé par une structure lumineuse qui spécifie ou renvoie de manière univoque à la surface arrière de l'objet.

Fodor et Pylyshyn se contentent d'affirmer que plusieurs expériences permettent d'établir que la structure lumineuse ne spécifie pas le « layout » (ou, autrement dit, n'y renvoie pas de manière univoque) et ils décrivent l'une de ces expériences. Pour se persuader de la valeur de l'affirmation de Fodor et Pylyshyn, il faudrait examiner l'ensemble de ces expériences. Ici, je me contenterai d'examiner l'expérience qu'ils ont décrite, et je vais tâcher de montrer que cette expérience n'établit pas l'équivocité ou ambiguïté de la structure optique. Plus précisément, je vais montrer que si l'expérience décrite par Fodor et Pylyshyn est interprétée de manière appropriée, nous devrons admettre ceci : la perception illusoire fait appel à des

inférences tandis que la perception non illusoire (normale ou véridique) ne s'appuie pas sur des inférences.

Cette expérience est relative, non pas à la perception visuelle, mais à la perception auditive. Il s'agit de faire entendre au sujet un mot au sein duquel une lettre est masquée par un toussotement. Ainsi, on fait entendre au sujet le mot « de#ermine », où le symbole « # » tient lieu du toussotement. Lorsque le sujet est soumis à ces sons, il entend la lettre « t » là où il y a un toussotement. La même expérience est répétée pour le mot « es#ape », et le sujet entend alors la lettre « k » là où il y a en fait un toussotement.

Selon l'analyse de Fodor et Pylyshyn, le sujet est soumis au même stimulus effectif, c'est-à-dire les ondes sonores causées par le toussotement, et il y a pourtant un « layout » différent pour chaque toussotement; dans une situation, le sujet perçoit un « t » et, dans l'autre situation, il perçoit un « k ». Le stimulus effectif manifeste donc de l'équivoque ou de l'ambiguïté relativement au « layout ». Or, la perception, elle, n'a pas un caractère équivoque. Le sujet entend ou bien un « t », ou bien un « k ». La perception comprend donc davantage d'information que le stimulus effectif. Donc, conformément à ce qui a été dit ci-dessus, cet excédent d'information au sein du contenu perceptuel doit être expliqué par l'intervention d'inférences.

Je ne crois pas que l'interprétation de cette expérience par Fodor et Pylyshyn soit valable. Il est faux de dire que le stimulus effectif est équivoque; qu'il renvoie à des « layouts » différents. Le « layout » est dans le monde; c'est ce qui est à l'origine du stimulus effectif. Or, la lettre « t » et la lettre « k », que Fodor et Pylyshyn considèrent comme des « layouts », ne sont nullement dans le monde et ils ne sont pas non plus à l'origine du stimulus effectif. Ce qui est dans le monde et qui est à l'origine du stimulus effectif (les ondes sonores), c'est le toussotement. C'est donc ce dernier qui est le « layout ». Or, le stimulus effectif est corrélé nomologiquement avec ce « layout »; en effet, les ondes sonores (le stimulus effectif) renvoient de manière univoque au toussotement. Donc, Fodor et Pylyshyn ont tort de poser que le stimulus effectif est équivoque. Doit-on conclure que les perceptions des lettres « t » et « k » sont non inférentielles; que ces perceptions peuvent se satisfaire de l'information contenue dans le stimulus effectif? Non, car l'information comprise dans ces perceptions est différente de l'information comprise dans le stimulus effectif. En effet, les perceptions ont

pour contenu les lettres « t » et « k », tandis que le contenu informationnel du stimulus effectif est un toussotement. Mais si le contenu informationnel de la perception diffère ainsi du contenu informationnel du stimulus effectif, elle ne peut pas avoir reçu son contenu informationnel du stimulus effectif. Il faut que son contenu provienne d'inférences. Cependant, ces perceptions, qui reçoivent leur contenu par le biais d'inférences, ne sont pas des perceptions véridiques ou normales; ce sont des perceptions *illusoires*. En effet, comme le contenu de ces perceptions diffère du « layout » (le toussement), il s'agit d'un contenu perceptuel *illusoire*. Plus précisément, comme j'entends les lettres « t » ou « k » là où il n'y a *en réalité* qu'un toussotement, ces perceptions auditives ne peuvent être qu'illusoires.

L'expérience décrite par Fodor et Pylyshyn nous montre donc que les perceptions illusoires gagnent leur contenu par un recours à des inférences. Mais cette expérience ne montre certainement pas que la perception normale a recours à des inférences. Dans la perception normale ou véridique, la perception est conforme au « layout ». Donc, pour montrer que la perception normale ou véridique a recours à des inférences, il faudrait pouvoir montrer que le stimulus effectif est équivoque relativement au « layout », et cela relativement à une perception normale. Or, l'expérience que Fodor et Pylyshyn ont décrite échoue sur ces deux points. D'une part, cette expérience ne porte pas sur une perception normale. D'autre part, cette expérience ne montre pas que le stimulus effectif est équivoque; au contraire, dans cette expérience, le stimulus effectif (les ondes sonores) spécifie le (renvoie de manière univoque au) « layout » (le toussotement).

4.4 Résumé

Dans la première partie du chapitre 4, nous avons examiné les 4 critères dont se sert Gibson pour circonscrire le domaine des propriétés directement perceptibles.

Il y a d'abord le critère phénoménologique. Fodor et Pylyshyn le rejettent pour trois raisons. Premièrement, ils font valoir que certaines propriétés jugées par Gibson comme étant directement perceptibles ne sont pas données à la conscience. Deuxièmement, ils font valoir que l'idée d'un donné phénoménologique est en soi une idée douteuse. Le donné phénoménologique suppose une grande quantité de données qui ne sont pas données à la conscience. Troisièmement, ils font valoir que l'identification des propriétés directement

perceptibles avec les propriétés phénoménologiques suppose la possibilité d'un mécanisme direct pouvant être à la base de la perception des propriétés phénoménologiques, en particulier des affordances. Or, un mécanisme direct pouvant servir à la perception des affordances semble être inconcevable. Je suis d'accord avec les deux premières raisons que Fodor et Pylyshyn invoquent pour rejeter le critère phénoménologique, mais pas avec la troisième. Le chapitre suivant va tenter d'élaborer un mécanisme direct pouvant détecter des invariants d'invariants, c'est-à-dire percevoir des affordances.

Le deuxième critère est le critère comportemental. Les propriétés directement perceptibles seraient les propriétés auxquelles répond le système perceptuel. Pour que ce critère soit valable, il faut pouvoir définir le système perceptuel. Or Gibson ne définirait pas adéquatement le système perceptuel. Il le définirait comme étant ce qui répond aux propriétés directement perceptibles, ce qui est la notion problématique du départ. Je suis d'accord avec cette critique de Fodor et Pylyshyn.

Le troisième critère est le critère écologique. Gibson définit les propriétés directement perceptibles comme étant des propriétés écologiques. Le problème, soutiennent Fodor et Pylyshyn, c'est que Gibson définit les propriétés écologiques comme des propriétés directement perceptibles, ce qui est la notion problématique du départ. Je ne suis pas d'accord avec cette critique de Fodor et Pylyshyn. Gibson ne se contente pas de définir les propriétés écologiques comme des propriétés directement perceptibles; il les définit également comme des propriétés qui sont situées à une échelle de grandeur significative pour les actions de l'agent. Les propriétés écologiques sont donc définies de manière adéquate, c'est-à-dire définies sans que l'on soit entraîné dans un cercle vicieux. Le critère écologique est donc valable pour définir les propriétés directement perceptibles. Cependant, ce critère est nécessaire, mais non suffisant. Toute propriété directement perceptible est une propriété située à une échelle de grandeur significative pour les actions de l'agent, mais il existe des propriétés situées à une échelle de grandeur significative pour les actions de l'agent qui ne sont pas directement perceptibles.

Le quatrième critère est nécessaire et suffisant et il englobe le précédent. C'est le critère écologique-nomologique. Il pose qu'une propriété directement perceptible est une propriété écologique qui réfléchit et structure nomologiquement la lumière. Ce critère est jugé par

Fodor et Pylyshyn comme étant non valable car il suppose que l'on puisse définir les lois de l'optique écologique et, selon Fodor et Pylyshyn, Gibson ne définit pas adéquatement les lois de l'optique écologique. Les lois de l'optique écologique sont des lois qui se rapportent aux propriétés écologiques, et, selon Fodor et Pylyshyn, ce genre de propriété est problématique puisqu'elles seraient définies par Gibson comme des propriétés directement perceptibles, la notion problématique du départ. Il n'est pas difficile de voir ce qui ne va pas dans cet argument. Les propriétés écologiques ne sont pas seulement définies par Gibson comme des propriétés directement perceptibles; elles sont aussi définies comme des propriétés qui se situent à une échelle de grandeur significative pour l'agent. La définition des lois de l'optique écologique n'est donc pas problématique. Le critère écologique-nomologique est donc valable; les propriétés directement perceptibles sont des propriétés écologiques qui réfléchissent et structurent nomologiquement la lumière.

Après avoir circonscrit le domaine des propriétés directement perceptibles, nous avons réfuté les 2 objections que Fodor et Pylyshyn adressent à la possibilité de la perception directe.

Dans leur première objection, Fodor et Pylyshyn acceptent la possibilité d'une détection directe de la structure optique, mais ils n'acceptent pas que la perception d'une propriété écologique (S_2) soit identifiée à la détection de la structure optique (S_1) qui est corrélée nomologiquement à la propriété écologique. Leur argument prend la forme suivante.

Selon la théorie de l'information, « S_1 contient de l'information sur S_2 » signifie seulement « S_1 est corrélé nomologiquement à S_2 ». Donc, S_1 ne *contient* pas en soi S_2. Or, si S_1 ne contient pas en soi S_2, la détection directe de S_1 n'est pas identifiable à la saisie du contenu « S_2 » (la perception de S_2).

Pour répondre à l'objection de Fodor et Pylyshyn, il a fallu compléter le modèle de Gibson. Pour qu'une structure indicative (pour nous : l'état du transducteur ou détecteur) acquiert comme contenu ou signification la propriété qu'elle indique (la propriété avec laquelle la structure indicative est corrélée nomologiquement), il faut que, en plus d'être corrélé nomologiquement avec cette propriété (en plus d'indiquer cette propriété), la structure indicative ait la fonction d'indiquer cette propriété. Donc, pour que l'état du transducteur ou détecteur ait S_2 comme contenu ou signification, il faut (et il suffit) que, en plus d'indiquer

S_2(1ère condition), l'état du détecteur ait la fonction d'indiquer S_2 (2ème condition). Or, via l'évolution ou l'apprentissage, l'état du détecteur peut recevoir cette fonction. Donc, lorsque ces deux conditions sont remplies, la détection de S_1 est effectivement identifiable à la saisie du contenu « S_2 » (la perception de S_2).

La deuxième objection de Fodor et Pylyshyn concerne la quantité d'information associée au stimulus effectif. Ils partent de ce constat : si le stimulus effectif contient moins d'information que l'information associée au « layout » ou, autrement dit, si le stimulus effectif est un signal équivoque relativement au « layout », alors il faut faire intervenir des inférences pour compléter l'information ou éliminer l'équivoque. À l'aide d'une expérience, ils tentent de montrer que le signal est effectivement équivoque relativement au « layout ».

J'ai montré que cette expérience n'établissait pas l'équivocité du stimulus effectif. Autrement dit, j'ai montré que le stimulus effectif (les ondes sonores) renvoyait de manière univoque au « layout » (le toussotement). Pourtant, même en présence d'un stimulus effectif univoque, le sujet perçoit deux sons différents (la perception du « t » et du « k »). Compte tenu que le stimulus effectif renvoie au même « layout » (le toussotement), comment expliquer ces deux perceptions différentes? Comme le stimulus effectif renvoie au même « layout », il porte l'information sur ce « layout ». Or, les deux perceptions n'ont pas ce « layout » comme contenu informationnel. Donc, le processus perceptuel ne s'en tient pas à l'information portée par le stimulus effectif; il fait intervenir un autre contenu informationnel. Cela est possible si le processus perceptuel fait intervenir des inférences. Ces deux perceptions sont donc vraisemblablement le fruit d'inférences. Devons-nous conclure que la perception en général est basée sur des inférences? Non, car ces deux perceptions sont des perceptions illusoires. En effet, ces deux perceptions ne correspondent pas au « layout ». Pour conclure que la perception normale ou véridique repose sur des inférences, il faudrait montrer que le stimulus effectif est équivoque relativement au « layout », et cela, Fodor et Pylyshyn ne le font pas. Donc, en définitive, cette expérience a montré que la perception illusoire repose sur des inférences, mais elle n'a pas montré que la perception normale ou véridique repose sur des inférences.

CHAPITRE V

LE RÉSEAU DE NEURONES À LA BASE DE LA PERCEPTION DES AFFORDANCES

Dans le chapitre précédent, j'ai montré que la perception d'une propriété écologique était identifiable à la détection de la structure optique. Fodor et Pylyshyn identifient « détection directe » et « transduction ». Je ne crois pas que la détection directe soit identifiable à un mécanisme de transduction. Selon moi, le mécanisme permettant de détecter la structure optique correspond à un réseau de neurones, plus précisément, au réseau de neurones récurrent simple de Elman (1990). C'est ce que je compte montrer au cours de ce chapitre. Plus précisément, je compte montrer que ce réseau de neurones est en mesure de détecter la structure optique qui spécifie les affordances. Comme on l'a vu au chapitre 3, la structure optique qui spécifie les affordances se caractérise comme un invariant d'invariants. Je compte donc montrer que le réseau récurrent simple de Elman est en mesure de détecter des invariant d'invariants.

Dans la première partie du chapitre 5, je vais présenter certaines des caractéristiques des réseaux de neurones. Cette présentation devrait être en mesure de nous faire comprendre les propos de la deuxième partie, où je définis un réseau de neurones qui est en mesure d'extraire des invariants d'invariants. Comme je viens de le dire, ce réseau de neurones correspond en grande partie au réseau de neurones récurrent simple de Elman (1990).

5.1. Une introduction partielle aux réseaux de neurones

Un réseau de neurones est un ensemble d'unités connectées entre elles. De manière plus ou moins analogue aux neurones, ces unités sont de petits systèmes computationnels. Donc, une unité reçoit un input et, sur la base de cet input, elle calcule un output. Dans beaucoup de modèles connexionnistes, mais pas dans tous, l'output d'une unité est égal à son activation.

Les connexions entres les unités sont pondérées. Ainsi, l'output de l'unité I aura plus ou moins d'influence sur l'unité U, selon que le poids de la connexion entre I et U est plus ou moins grand. Le produit de l'output par le poids de la connexion ($output_i \times poids_{ui}$) correspond à l'input de l'unité U. Si l'unité U est connectée à d'autres unités, l'unité U reçoit des inputs de ces autres unités, et l'input total de l'unité U correspond à la somme de tous ces inputs partiels. L'activation de l'unité U est fonction de l'input total. La règle d'activation la plus simple pose que l'activation d'une unité est égale à son input total. Pour une autre règle d'activation, l'activation peut être 1 ou 0. Elle est de 1 lorsque l'input total atteint ou dépasse une valeur seuil (threshold) et elle est de 0 lorsque la valeur de l'input total est inférieure au seuil. Il existe une autre règle d'activation : il s'agit de la règle logistique. On la calcule de cette façon : $a_u = (1+ e^{-netinput})^{-1}$.

Les réseaux de neurones qui nous intéressent sont des réseaux de neurones « feedforward ». Les réseaux « feedforward » contiennent minimalement deux couches d'unités[45]. La première couche, la couche d'input, est connectée unidirectionnellement à la deuxième couche, la couche d'output. Autrement dit, seules les unités de la couche d'output reçoivent des inputs des unités de la couche d'input. Habituellement, mais pas dans tous les modèles, chaque unité de la couche d'output reçoit un input de toutes les unités de la couche d'input.

Comment l'activation dans un réseau « feedforward » à deux couches se propage-t-elle? D'abord, les unités de la couche d'input reçoivent leur activation d'un input externe. Le *pattern d'activation* de la couche d'input, c'est-à-dire le vecteur qui représente l'activation que prend chacune des unités de la couche d'input, est identique au pattern d'activation de l'input externe. Ensuite, chacune des unités de la couche d'input émet un output qui, dans la plupart des modèles, est identique à la valeur de l'activation. Cet output, une fois qu'il est multiplié par le poids de la connexion, tient lieu d'input pour l'unité de la couche d'output. Donc, si une unité de la couche d'output est connectée à plusieurs unités d'input, l'unité de la couche d'output reçoit plusieurs inputs. L'input total (netinput) de l'unité d'output est égal à la sommation de tous ces inputs. Après avoir reçu leur input total, chaque unité de la couche d'output calcule la valeur de son activation sur la base de l'une ou l'autre des règles

[45] Chaque couche contient une ou plusieurs unités.

d'activation présentées plus haut. Ainsi, lorsque l'activation de chacune des unités de la couche d'output est considérée, la couche d'output acquiert un pattern d'activation.

Suivant l'interprétation de celui qui a conçu le réseau de neurones, le pattern d'activation de la couche d'output et de la couche d'input représente quelque chose de particulier. Si la tâche du réseau de neurones est de classifier des objets, le pattern d'activation de la couche d'output représente une classe particulière, tandis que le pattern d'activation de la couche d'input représente un objet à classifier. L'objectif du concepteur est bien simple : lorsque le pattern d'activation de la couche d'input représente un objet particulier, la propagation de l'activation au sein du réseau de neurones doit causer au sein de la couche d'output un pattern d'activation qui représente la classe à laquelle appartient cet objet. Le concepteur parvient à ses fins en modifiant le poids des connexions. En effet, la façon dont le pattern d'activation de la couche d'input se propage au sein du réseau est fonction du poids des connexions. Le changement du poids des connexions se produit lors de l'apprentissage. Les réseaux de neurones peuvent faire l'objet de plusieurs types d'apprentissage. Cependant, dans le cadre du présent chapitre, je ne traiterai pas de ces différentes formes d'apprentissage.

Jusqu'ici, je n'ai traité que des réseaux « feedforward » à deux couches. La capacité d'un réseau à deux couches est cependant limitée; pour accomplir leur tâche, les réseaux de neurones « feedforward » doivent souvent insérer une ou plusieurs couches entre les couches d'input et d'output. Ces couches intermédiaires sont appelées « couches cachées ».

Le réseau à deux couches est incapable d'accomplir certaines tâches car l'information contenue dans le pattern d'activation de la couche d'input rend impossible son traitement en une seule étape[46]. Autrement dit, l'information de la couche d'input ne peut pas être traitée par la couche d'output de manière à ce que cette dernière reçoive le pattern d'activation voulu. Un traitement intermédiaire est nécessaire; la couche d'output peut acquérir le bon pattern d'activation seulement si elle traite une information qui a déjà été traitée lors d'une étape préalable, et c'est la couche cachée qui exécute ce traitement préalable.

---

[46] Cela se produit lorsque les paires input-output ne sont pas linéairement séparables (Elman *et al.*, 1998). La plupart des tâches font intervenir des paires input-output qui ne sont pas linéairement séparables.

> In order to solve problems like this, it is necessary to insert hidden units between input and output units that can preprocess the information to obtain the pieces that are needed for the final solution of the overall problem. This produces a more complex flow of information in which intermediate results [...] are obtained by combining input activations in certain ways; it is then the final set of intermediate results that is used by the output units. (Bechtel et Abrahamsen, 2002)

Un exemple pourrait donner une valeur plus intuitive à ce qui vient d'être dit à propos de la nécessité d'introduire une ou des couches cachées. Supposons que la tâche d'un réseau de neurones est de reconnaître des formes. Les unités de la couche d'input correspondent à des capteurs sensoriels d'ordre visuel. Plus précisément, le pattern d'activation de la couche d'input correspond à une distribution de points d'intensité variable. Comme le montre les analyses portant sur l'excitation des cônes dans l'œil, cette distribution de points d'intensité variable peut être « brouillée ». Donc, à partir des patterns d'activation situés sur la couche d'input, le réseau de neurones peut être incapable de sortir le pattern d'output qui représente la forme de l'objet. Un traitement supplémentaire de l'information est requis. Ainsi, une couche cachée pourrait traiter cette information et, à la suite de ce traitement, elle pourrait acquérir un pattern d'activation qui représente les bords (edges) de l'objet. Et, maintenant, sur la base de cette nouvelle information, c'est-à-dire une information qui nous renseigne sur les bords de l'objet, la couche d'output peut être activée de manière à acquérir un pattern d'activation qui représente la forme de l'objet. Comme on le voit, la représentation (le pattern d'activation) de la couche cachée est une représentation « améliorée » de la représentation contenue dans la couche d'input. Cette représentation est « améliorée » eu égard à la fonction de la couche d'output, qui est de trouver la forme de l'objet.

Le réseau de neurones qui est à la base de la perception des affordances est un réseau à couche cachée. Cependant, il ne s'agit pas d'un réseau exclusivement « feedforward ». Il s'agit d'un réseau de neurones qui possède des connexions récurrentes. Plus précisément, on appelle ces réseaux « réseaux récurrents simples ». On retrouve deux types majeurs de réseaux récurrents simples, celui de Elman (1990) et celui de Jordan (1986). Je vais les présenter brièvement.

Le réseau récurrent simple de Elman a tout d'un réseau « feedforward » à couche cachée, sauf que des connexions partent de la couche cachée vers une couche d'unités, appelée

« couche contextuelle », qui se situe au niveau de la couche d'input et qui, comme la couche d'input, alimente la couche cachée. Le réseau de Jordan ressemble au réseau de Elman, sauf que les connexions récurrentes partent de la couche d'output au lieu de partir de la couche cachée. La couche d'unités que ces connexions récurrentes alimentent est appelée par Jordan « couche d'état ». Aussi, Jordan introduit dans son réseau des connexions récurrentes qui partent des unités de la couche d'état et qui retournent sur elles-mêmes. Aussi, il faut souligner que les connexions récurrentes ne modifient pas le pattern d'activation de la couche d'où elles partent : le poids des connexions récurrentes est de 1. Les connexions récurrentes permettent donc de copier un pattern d'activation sur une autre couche (la couche contextuelle ou d'état). Les deux figures suivantes illustrent les modèles de Jordan et de Elman.

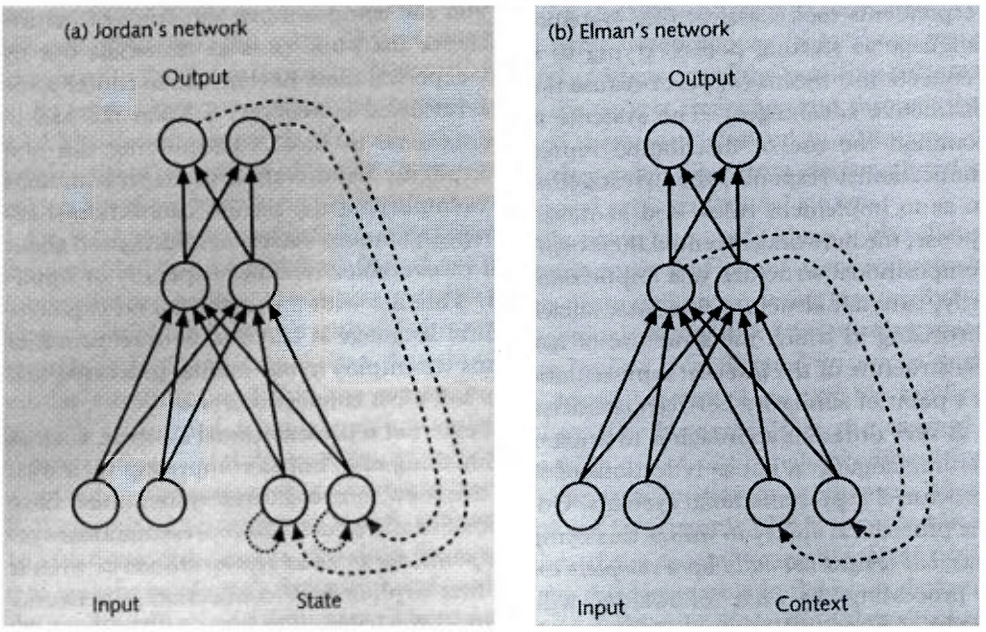


Figure 5.1 Les réseaux de Elman et Jordan.

Le mécanisme permettant la détection d'invariants prendra la forme de l'un de ces deux réseaux de neurones. Plus loin dans ce chapitre, je serai donc amené à choisir entre le réseau de Elman et le réseau de Jordan. Comme je l'ai déjà dit, je choisirai le réseau de Elman.

5.2. Le réseau de neurones à la base de la perception des affordances

Le réseau de neurones qui nous intéresse doit pouvoir représenter sur sa couche d'output un invariant d'invariants. Avant de nous pencher sur un réseau de neurones qui soit en mesure de représenter un invariant d'invariant sur sa couche d'output, nous allons tenter de trouver un réseau de neurones qui soit en mesure de représenter un invariant sur sa couche d'output.

Comme on l'a vu, l'invariant est ce qui est commun à une succession de structures de perspectives. Donc, l'input de notre réseau de neurones est une succession de structures de perspectives. Une fois qu'il est soumis à cet input temporel, le réseau de neurones doit sortir sur sa couche d'output l'invariant correspondant à cette succession d'input. Mais, si l'input du réseau de neurones est une succession de structures de perspectives, le réseau doit mémoriser des structures de perspectives, puisqu'une succession implique évidemment des moments antérieurs. Donc, dans notre recherche d'un réseau de neurones pouvant extraire des invariants, le premier problème qui se dresse est celui de trouver un réseau qui peut mémoriser des structures de perspectives. Les réseaux de Elman (1990) et Jordan (1986) permettent cela.

5.2.1 Un réseau de neurones doté d'une mémoire dynamique

Traditionnellement, les psychologues ont situé la mémoire *statique* du réseau de neurones dans le poids des connexions. Une fois que le réseau a été entraîné, le poids des connexions ne change pas; ce que le réseau a appris durant son entraînement reste toujours identique — d'où l'expression « mémoire statique ». Ce n'est pas ce type de mémoire que je recherche maintenant. Je suis à la recherche d'une mémoire dynamique, c'est-à-dire une mémoire qui retient chaque *nouvel* input et qui, de ce fait, est en *changement* continuel.

Les connexions récurrentes des réseaux de Elman et de Jordan permettent d'introduire une mémoire dynamique au sein du réseau de neurones. Pour montrer cela, je vais recourir aux simulations que Elman a produites dans son article *Finding Structure in Time* (1990).

Le premier problème auquel Elman a soumis son réseau est le problème du OU exclusif ou, plus simplement, le problème « XOR ». D'habitude, ce problème est traité de manière

statique. Le réseau est soumis à des inputs de deux bits (01, 10, 00, 11) et le réseau sort sur sa couche d'output des vecteurs de 1 bit (1, 1, 0, 0)[47]. On peut donner une dimension temporelle à ce problème. On crée une séquence d'inputs de 1 bit où les deux premiers inputs correspondent à l'input de deux bits et où le troisième input de la séquence correspond à l'output des deux bits précédents. Et ce cycle est répété, en prenant bien soin de choisir au hasard les paires d'input-output. On pourrait ainsi obtenir la séquence suivante :

000101110011101 ...

La tâche du réseau de Elman est de prédire le prochain input. On entraîne d'abord le réseau en lui soumettant 600 fois une séquence de 3000 bits et, vraisemblablement, en usant de la méthode de rétropropagation[48].

Le réseau ne peut pas prédire chaque input, puisque, sauf lorsqu'il est soumis aux deux inputs successifs qui « décompose » temporellement l'input de deux bits, les chances de bien prédire le prochain input sont de 50%. Ainsi, lorsqu'il est soumis au premier input, les chances de prédire correctement le prochain input sont de 50%, tandis que s'il est soumis au deuxième input, le réseau devrait pouvoir prédire le troisième input, puisque les deux premiers inputs suffisent pour conclure à propos du troisième input. Une fois qu'il est soumis au troisième input, ces chances de prédire le prochain input sont encore de 50%; idem pour le quatrième; tandis qu'il devrait prédire correctement le prochain input après avoir été soumis au cinquième input, puisque le quatrième et le cinquième input suffisent pour porter une conclusion à propos du sixième input. Et ce cycle devrait se répéter. C'est effectivement ce qui se passe lorsque, après avoir entraîné son réseau, Elman le teste. Voici une figure qui indique l'erreur que le réseau de Elman produit lorsqu'il est testé sur une période de temps qui s'étend sur 12 inputs successifs.

---

[47] Le réseau répond positivement (1 ou vrai) seulement quand un bit est positif (1 ou vrai) et quand l'autre est nul (0 ou faux). Lorsque les deux bits sont identiques, le réseau répond « 0 ou faux ».

[48] Dans son article de 1990, Elman ne spécifie la méthode d'apprentissage qu'il utilise. Cependant, comme il utilise un réseau à couche cachée, on peut présumer qu'il s'agit de la méthode de rétropropagation. Mais il pourrait également s'appuyer sur un apprentissage par renforcement.

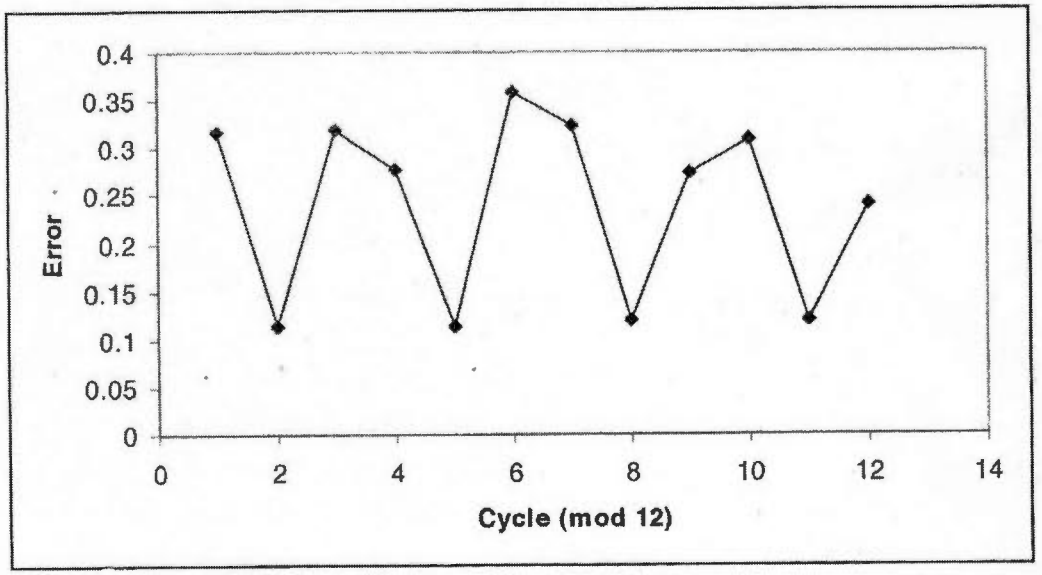


Figure 5.2 L'erreur associée à l'output. Tirée de Elman (1990).

Donc, lorsqu'il est possible de prédire le prochain input, le réseau parvient effectivement à le prédire. Mais pour prédire ainsi le prochain input, le réseau doit être en mesure de mémoriser l'input précédent. En effet, lorsque, sur la base des $2^{ème}$, $5^{ème}$, $8^{ème}$ et $11^{ème}$ inputs, le réseau prédit ce que sera le prochain input, il ne se fonde pas seulement sur ces inputs singuliers; il tient compte de l'input qui précède ces inputs. En effet, c'est sur la base de l'input précédent et de l'input actuel, et non pas seulement sur la base de l'input actuel, que l'on peut prédire le prochain input. En fait, ce qui se passe au niveau du réseau de neurones est assez simple.

Prenons par exemple les deux premiers inputs. Le réseau reçoit d'abord le premier input. Le pattern d'activation de la couche d'input active ensuite la couche cachée. À son tour, le pattern d'activation de la couche cachée active la couche d'output ainsi que la couche contextuelle. Les connexions allant de la couche cachée vers la couche contextuelle possèdent une valeur de 1. Le pattern d'activation de la couche contextuelle est donc une copie du pattern d'activation de la couche cachée. Le pattern d'activation de la couche contextuelle (ou la copie de la couche cachée) agira au prochain cycle. Donc, au prochain cycle, la couche contextuelle aura retenu ou mémorisé le pattern d'activation de la couche

cachée qui avait lieu au cycle précédent. Et comme le pattern d'activation de la couche cachée est une représentation « améliorée » de la couche d'input, nous pouvons dire que la couche contextuelle retient ou mémorise le premier input. Ainsi, si l'on revient au problème « XOR », on peut supposer que, lorsque le second input survient, la couche cachée reçoit de l'information à la fois de l'input actuel et de l'input précédent, si bien que, lorsque la couche d'output prédit le prochain input, elle se base sur ces deux inputs et non pas seulement sur l'input actuel.

Grâce à ces connexions récurrentes, le réseau de Elman peut donc retenir ou mémoriser l'input précédent. Mais tout cela est-il suffisant? La couche d'input impliquée dans la détection d'invariants ne peut pas être un vecteur possédant seulement une dimension ou 1 bit; or, la simulation présentée dans l'article de Elman fait intervenir des inputs ayant seulement un bit (ou une dimension). En outre, la détection d'invariants se base sur des inputs qui reculent assez loin dans le temps; or, la simulation de Elman fait intervenir une mémoire qui recule simplement d'un pas vers l'arrière. Donc, pour se persuader du fait que la mémoire du réseau de Elman est en mesure de fonder la détection d'invariants, il faudrait montrer que ce réseau mémorise des inputs ayant plus d'une dimension (plus de 1 bit) et il faudrait montrer que cette mémoire recule plus loin en arrière. La seconde simulation de Elman, toujours présentée dans *Finding structure in time* (1990), permet de montrer cela.

La seconde simulation en est toujours une de prédiction. Les séquences sont composées de lettres et les lettres sont représentées par des vecteurs binaires à 6 dimensions, au lieu d'une seule, comme dans la simulation précédente. Il y a trois séquences possibles et chacune de ces séquences commence par une consonne suivie de voyelles. Les trois séquences sont les suivantes : *ba*, *dii* et *guuu*. Ces trois séquences sont combinées au hasard pour former une longue séquence de 1000 lettres. Cette séquence de 1000 lettres (ou vecteurs à 6 dimensions) est soumise au réseau 200 fois en vue de lui apprendre à prédire la prochaine lettre. Bien qu'Elman ne le spécifie pas, Elman procède vraisemblablement par rétropropagation pour entraîner son réseau. Après cet entraînement, les capacités prédictives du réseau sont testées sur une autre séquence, où les mêmes régularités sont observées. Les seules lettres (vecteurs ou inputs) que le réseau ne devrait pas être capable de prédire sont les consonnes et il devrait

être capable de prédire les voyelles. C'est précisément ce qui se passe. La figure suivante, tirée de Elman (1990), l'atteste.

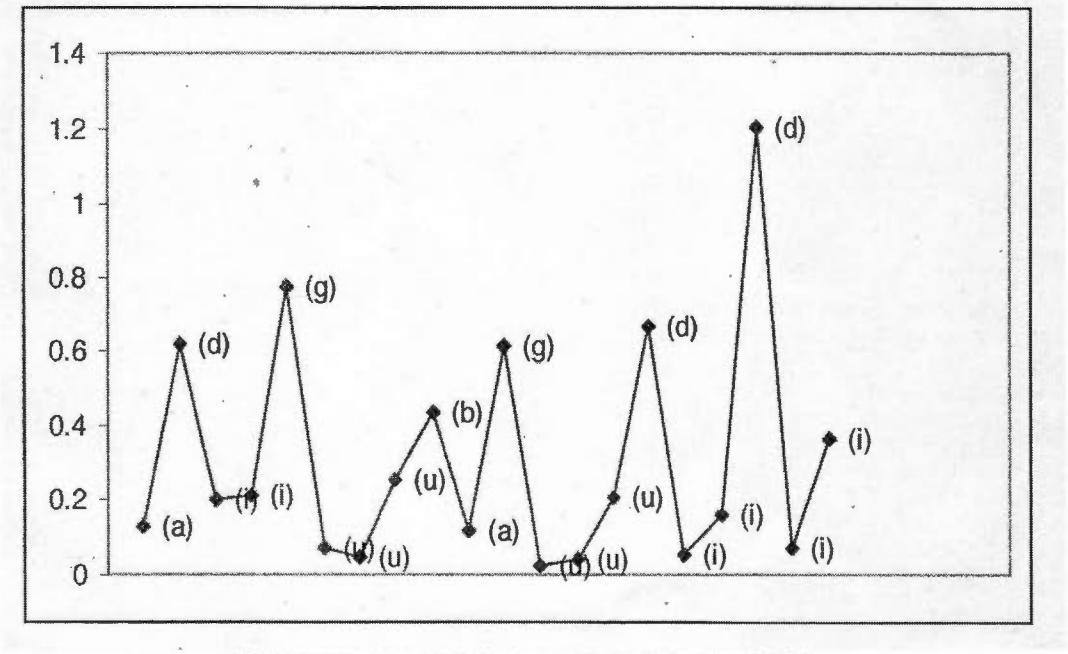


Figure 5.3 L'erreur associée à l'output. Tirée de Elman (1990).

En fait, le réseau prédit beaucoup plus que les voyelles. Il est également en mesure de prédire qu'après la succession des voyelles, la lettre qui suivra sera une consonne. C'est ce que l'on observe si l'on tient compte de l'erreur associée à la première dimension des vecteurs représentant les lettres. Comme l'indique le tableau ci-dessous, la première dimension spécifie si la lettre est une consonne et Elman a constaté que l'erreur associée à cette première dimension était très basse lorsqu'il s'agissait de prédire la lettre (une consonne) qui suit la séquence de voyelles.

Tableau 5.1 Les vecteurs représentant les lettres. Tiré de Elman (1990).

	Consonant	Vowel	Interrupted	High	Back	Voiced
b	[1	0	1	0	0	1]
d	[1	0	1	1	0	1]
g	[1	0	1	0	1	1]
a	[0	1	0	0	1	1]
i	[0	1	0	1	0	1]
u	[0	1	0	1	1	1]

Manifestement, pour prédire ainsi les voyelles, surtout les voyelles qui concernent la séquence *guuu*, le réseau doit retenir ou mémoriser davantage que l'input précédent. En fait, lorsqu'il s'agit de prédire la consonne qui suit les trois voyelles de *guuu*, le réseau doit pouvoir retenir ou mémoriser trois inputs antérieurs. Donc, apparemment, le réseau peut mémoriser un nombre considérable d'inputs successifs.

Maintenant que la capacité de mémorisation des réseaux récurrents simples apparaît suffisante pour être appliquée à la détection des invariants, il est temps de rendre vraisemblable la thèse selon laquelle un réseau récurrent simple peut détecter des invariants.

5.2.2. Jordan ou Elman?

En lisant le texte de Jordan (1986), on peut être tenté de penser que son réseau récurrent simple n'est pas en mesure d'extraire ou de détecter ce qui est invariant au sein d'une succession d'inputs. Dans le texte de Jordan, la variabilité de l'output du réseau ne vient pas de la couche d'input, mais de la couche d'état. En effet, l'input est statique; plus précisément, l'input consiste en un plan statique. Or, la détection d'un invariant fait référence non pas à un input qui reste identique (un input statique), mais à une succession d'inputs différents. Dans ces circonstances, on peut effectivement s'interroger sur la capacité du réseau de Jordan à détecter des invariants. Je ne crois pas que l'on puisse critiquer le réseau de Jordan sur cet aspect. En fait, mon point est le suivant : ce n'est pas parce que le texte de Jordan fait

intervenir un input statique que le réseau de neurones de Jordan est incapable de gérer un input dynamique; il apparaît possible de le soumettre à un input dynamique, c'est-à-dire à une succession d'inputs différents. Dans ce qui suit, nous allons donc supposer que le réseau de Jordan peut être soumis à un input dynamique. C'est sur un autre aspect que je vais critiquer ce réseau. Mais avant de montrer que le réseau de Jordan ne fait pas l'affaire, montrons que, en principe, le réseau de Elman est capable de faire le travail qui nous intéresse, c'est-à-dire de détecter des invariants.

5.2.2.1. Comment le réseau de Elman permet de détecter des invariants?

Tout d'abord, la couche d'input du réseau de Elman est activée par une structure de perspectives. L'activation de la couche d'input (son pattern d'activation) alimente ensuite la couche cachée. L'activation de la couche cachée est une représentation « améliorée » de la couche d'input. Cette représentation alimente ensuite la couche d'output et la couche contextuelle. Compte tenu que les connexions allant de la couche cachée vers la couche contextuelle possèdent toutes une valeur de 1, la couche contextuelle est une copie de l'activation de la couche cachée, c'est-à-dire une représentation « améliorée » de l'input initial. Ensuite, le second input survient. Maintenant, l'activation de la couche cachée résulte de la couche d'input et de la couche contextuelle, où le premier input a été mémorisé. Ces deux couches d'unités — la couche d'input et la couche contextuelle — sont connectées à la couche cachée de manière à ce que l'activation de la couche cachée soit une représentation des deux inputs qui puisse faire l'objet d'un traitement permettant de détecter ce qui est invariant au sein de ces deux inputs. Ce traitement a lieu sur la couche d'output, à laquelle est connectée la couche cachée. C'est donc l'activation de la couche d'output qui représente l'invariant des deux premiers inputs. Mais la couche cachée alimente aussi la couche contextuelle. En fait, comme on l'a souligné, l'activation de la couche contextuelle est une copie de l'activation de la couche cachée. Donc, sur la couche contextuelle, on retrouve une copie des deux inputs, puisque la couche cachée représentait de manière « améliorée » ces deux inputs. Ensuite survient le troisième input. La couche d'input est activée une nouvelle fois. La couche cachée sera donc activée par la couche d'input et par la couche contextuelle, qui mémorise les deux inputs précédents. La couche cachée, lorsqu'elle sera activée par ces deux couches, représentera donc de manière « améliorée » les trois inputs. Et, de manière

similaire à ce qui s'est passé plus tôt, cette représentation « améliorée » des trois inputs doit pouvoir être traitée de manière à ce qu'on puisse en tirer un invariant. Ce traitement va se réaliser sur la couche à laquelle est connectée la couche cachée, c'est-à-dire sur la couche d'output. C'est donc la couche d'output qui représente l'invariant des trois premiers inputs. Et ce cycle se répète lorsque surviennent de nouveaux inputs. C'est ainsi que l'on parvient à détecter des invariants.

Donc, apparemment, le réseau de Elman devrait être en mesure de détecter des invariants. Cependant, pour en être absolument certain, il faudrait évidemment entraîner un réseau de Elman et, ensuite, le tester. Je ne produirai pas cette simulation ici. Par contre, je vais prendre la peine de décrire le genre d'entraînement auquel ce réseau devrait être soumis pour qu'il puisse détecter des invariants.

On devrait d'abord former des ensembles d'inputs et d'outputs désirés. Dans chaque ensemble, on retrouverait un input temporel, qui serait en fait formé d'au moins deux inputs successifs. À chaque fois que le réseau serait soumis à deux inputs successifs, par exemple « 011110 » et « 101101 », le réseau émettrait un output (un pattern d'activation sur la couche d'output) et cet output serait comparé à l'output désiré (001100), qui est l'invariant des deux inputs. On calculerait ensuite l'écart entre l'output actuel et l'output désiré (l'invariant) et, par rétropropagation, on ajusterait le poids des connexions pour que, la prochaine fois, le réseau émette l'output désiré (l'invariant) lorsqu'il sera soumis à ces deux inputs. C'est en créant une multitude d'ensembles d'input-output de ce genre et en passant ces ensembles d'input-output à travers le réseau un très grand nombre de fois que le réseau finirait par apprendre à détecter ce qui est invariant au sein d'inputs successifs.

Donc, en principe, le réseau de Elman est en mesure de détecter des invariants. Mais, comme le réseau de Jordan est aussi un réseau récurrent simple, pourquoi celui-ci ne serait-il pas capable de détecter des invariants? Je vais donner quelques raisons qui m'empêchent de croire que le réseau de Jordan soit en mesure de détecter des invariants et, ensuite, je vais tenter de trouver un mécanisme (un réseau de neurones) qui permet de détecter des invariants d'invariants.

5.2.2.2. Pourquoi le réseau de Jordan ne permet pas de détecter des invariants?

Il s'agit donc de montrer que, lorsqu'il est soumis à un input dynamique (c'est-à-dire à une succession d'inputs ou de structures de perspectives), le réseau de Jordan ne permet pas de détecter ce qui est invariant au sein de ces inputs successifs.

Supposons d'abord que le réseau ait déjà été soumis à plus d'un input (ou à plus d'une structure de perspectives). Cela signifie que le réseau a déjà eu le temps de calculer un ou plusieurs invariants, selon que les inputs auxquels le réseau a été soumis sont au nombre de deux ou plus. Supposons que le réseau a été soumis à trois inputs et que, donc, il a pu calculer deux invariants, un pour les deux premiers inputs et un deuxième pour les trois. Un quatrième input survient et le réseau doit calculer l'invariant se rapportant aux quatre inputs. Sur la couche d'état, qui équivaut à la couche contextuelle du réseau de Elman, l'information correspondant à la couche d'output est copiée. Donc, puisque la couche d'output calcule l'invariant, les deux invariants sont représentés dans la couche d'état. En effet, ce n'est pas seulement le dernier invariant qui est copié sur la couche d'état, mais c'est aussi l'information qui a été calculée antérieurement sur la couche d'output, c'est-à-dire l'information correspondant au premier invariant (celui qui se rapporte aux deux premiers inputs). Tout comme la couche contextuelle du réseau de Elman, qui mémorise les états antérieurs de la couche cachée, la couche d'état du réseau de Jordan mémorise aussi des états antérieurs, sauf qu'il s'agit des états antérieurs de la couche d'output[49], c'est-à-dire des invariants, puisque nous posons que le réseau calcule des invariants. Donc, lorsque le quatrième input survient, la couche cachée doit traiter l'information correspondant à l'input actuel et l'information correspondant aux deux invariants passés. Supposons donc que cette information soit traitée de manière à ce qu'une représentation de toute cette information soit calculée. Ensuite, la couche d'output doit traiter l'information provenant de la couche cachée de manière à représenter ce qui est invariant au sein des deux premiers invariants et du quatrième input. Que la couche d'output soit en mesure de calculer cet invariant sur la base

[49] « [T]he state [la couche d'état] can be defined as an exponentially average of past outputs, so that the arbitrarily distant past has some representation in the state [dans la couche d'état], albeit with ever-diminishing strength. » (Jordan, 1986, p. 536)

de l'information provenant de la couche cachée est parfaitement concevable. Il semble donc qu'à première vue, le réseau de Jordan soit capable de détecter des invariants.

On se convainc de l'incapacité du réseau de Jordan à calculer des invariants lorsque l'on tient compte du premier input, qui a été délibérément écarté de la description qui vient d'être donnée.

Supposons que le réseau est soumis à un premier input. L'information correspondant à la couche d'input est ensuite acheminée vers la couche cachée, qui calcule une représentation « améliorée » de l'input. Par la suite, l'information contenue dans la couche cachée est acheminée vers la couche d'output. La couche d'output a pour fonction de calculer ce qui est invariant dans l'information provenant de la couche cachée. Or, en l'occurrence, la couche cachée ne contient que l'information correspondant à une seule structure de perspectives. Il n'y a donc pas d'invariant qui puisse être tiré de cette information. Donc, sur la base de l'information venant de la couche cachée, la couche d'output va calculer un « invariant » biaisé.

À première vue, que le premier output, celui qui correspond au premier input, ne soit pas une réponse adéquate, est parfaitement normal, puisque, sur la base d'un seul input ou structure de perspectives, on ne peut pas tirer un invariant. Mais le problème du réseau de Jordan, c'est que, via les connexions récurrentes, il réinjecte dans le réseau l'erreur qui vient de se produire sur la couche d'output. Ceci rend impossible la détection d'invariant. On peut s'en convaincre avec ce qui suit.

Cet invariant biaisé est ensuite acheminé vers la couche d'état. Un deuxième input survient alors et l'information correspondant à cet input ainsi que celle qui correspond à l'invariant biaisé sont acheminées vers la couche cachée. La couche cachée a pour fonction de calculer une représentation de cette information. Ensuite, cette information doit être envoyée à la couche d'output, qui a pour fonction de calculer ce qui est invariant dans l'information provenant de la couche cachée. Or, étant donné le caractère biaisé de l'invariant provenant de la couche d'état, l'information de la couche cachée sera elle aussi biaisée. Donc, de même, l'invariant qui sera calculé sur la couche d'output sera aussi biaisé. Il ne sera pas possible de détecter l'invariant qui correspond aux deux premiers inputs. Et, comme ce nouvel invariant biaisé est réintroduit dans le réseau en raison des connexions récurrentes, il

ne sera pas davantage possible de détecter l'invariant qui tient compte de nouveaux inputs. Apparemment, donc, le réseau de Jordan ne permet pas de détecter des invariants.

5.2.3. La détection d'invariant d'invariants

C'est donc le réseau de Elman qui devrait nous permettre de détecter des invariants. Qu'en est-il maintenant de la détection d'invariant d'invariants?

Un invariant d'invariants, comme on l'a vu dans un chapitre précédent, est tout simplement une combinaison invariante d'invariants. Ainsi, l'affordance présentée par une chaise, la possibilité de s'asseoir, est une combinaison invariante des 5 invariants correspondant à 1) la hauteur de la surface horizontale, 2) la rigidité de la surface horizontale, et 3) la grandeur de la surface horizontale, etc. Le réseau de neurones permettant de détecter une affordance (un invariant d'invariants) doit donc d'abord comporter des réseaux de neurones pouvant détecter ces invariants. Ainsi, comme les invariants de premier niveau sont donnés par des réseaux de Elman, ce réseau de neurones doit comprendre comme sous-systèmes cinq réseaux de Elman, un pour chaque invariant de premier niveau.

Ces réseaux de Elman servent ensuite d'inputs à un réseau supérieur. Plus précisément, ce réseau supérieur est un réseau à deux couches ou à couche cachée dont la couche d'inputs correspond aux couches d'outputs des cinq réseaux de Elman. Je suis présentement dans l'impossibilité de dire si le réseau supérieur est un réseau à deux couches ou un réseau à couche cachée. Pour savoir cela, il faudrait déterminer au préalable si les paires input-output sont linéairement séparables.

De plus, ce réseau supérieur (à deux ou trois couches?) n'a pas besoin de connexions récurrentes; il s'agit simplement d'un réseau « feedforward ». Ce réseau supérieur n'a pas besoin de connexions récurrentes car l'output (l'invariant d'invariants) de ce réseau n'est pas fonction d'un input (l'ensemble des outputs des réseaux de Elman) dynamique. En effet, pour que le réseau supérieur émette son output, les invariants de premier niveau doivent être présent simultanément, et non pas successivement.

5.3 Résumé

L'objectif du chapitre 4 était de trouver un réseau de neurones qui puisse détecter des invariants d'invariants. Afin d'atteindre cet objectif, nous avons d'abord essayé de trouver un réseau de neurones qui puisse saisir ou détecter des invariants.

La caractéristique fondamentale d'un réseau qui doit pouvoir détecter des invariants est de pouvoir mémoriser les inputs. Autrement dit, ce réseau doit posséder une mémoire dynamique. Ce réseau doit posséder une mémoire dynamique car l'invariant correspond à ce qui est commun ou invariant au sein d'un ensemble de structures de perspectives *successives*, et, pour pouvoir extraire ce qui est invariant au sein d'une succession de structures de perspectives, il faut retenir ou mémoriser les structures de perspectives antérieures.

Les réseaux récurrents simples permettent de mémoriser les états antérieurs du réseau. À l'aide de deux simulations tirées de Elman (1990), j'ai montré que cela était effectivement le cas. Plus précisément, Elman montre que son réseau peut prédire le prochain input à partir de l'input actuel et de (des) l'input(s) antérieur(s). Fort de ce résultat, j'ai ensuite montré que le réseau de Elman était en mesure de détecter des invariants.

Les choses se passent à peu près ainsi. Tout d'abord, une structure de perspectives vient activer la couche d'input. Cette activation se propage ensuite vers la couche cachée, où prend place une représentation « améliorée » de la structure de perspectives. Ensuite, l'activation de la couche cachée se propage vers la couche d'output et la couche contextuelle. L'activation de la couche contextuelle est une copie de l'activation de la couche cachée. Le prochain cycle commence alors; une nouvelle structure de perspectives active la couche d'input. L'activation de la couche d'input ainsi que celle de la couche contextuelle se propage alors vers la couche cachée. L'activation de la couche cachée représente l'input actuel (venant de la couche d'input) et l'input antérieur (venant de la couche contextuelle). La représentation qui a lieu sur la couche cachée va être traitée par la couche d'output et cette représentation est telle que la couche d'output va pouvoir trouver ce qui est invariant au sein des deux inputs.

Au moment où l'activation de la couche cachée se propage vers la couche d'output, une copie de l'activation de la couche cachée est envoyée vers la couche contextuelle. Donc, la couche contextuelle contient une représentation des deux premiers inputs. Ensuite, le

deuxième cycle commence. La couche d'input est activée une nouvelle fois par une troisième structure de perspectives. La couche d'input et la couche contextuelle activent alors la couche cachée. La couche cachée représente les trois inputs et cette représentation est telle que la couche d'output va pouvoir trouver ce qui est invariant au sein de ces trois inputs. Un deuxième invariant est trouvé. Les cycles pourraient se succéder ainsi et, pour chacun d'eux, un nouvel invariant pourrait être représenté sur la couche d'output.

Après avoir montré que le réseau de Elman est vraisemblablement en mesure de détecter des invariants, j'ai montré que le réseau de Jordan, un autre réseau récurrent simple, en était incapable. La couche d'output du réseau de Jordan traite le premier input (la représentation « améliorée » de la couche cachée) de manière à trouver ce qui est invariant en lui. Or, comme cet input est singulier, il ne possède pas d'invariant. La représentation qui aura lieu sur la couche d'output sera donc biaisée ou erronée. Le problème du réseau de Jordan, c'est que, par le biais de ses connexions récurrentes, il réinjecte cette erreur dans le réseau, et cela, comme je l'ai montré, rend impossible l'extraction d'un invariant.

En dernier lieu, j'ai esquissé la forme que doit prendre un réseau de neurones si celui-ci est en mesure de détecter des invariants d'invariants. Ce réseau comprend deux niveaux. Au premier niveau, nous retrouvons des réseaux de Elman. Comme on l'a vu, l'affordance est constituée d'une combinaison invariante de propriétés occurrentes. À chaque propriété occurrente de l'affordance, il faut associer un réseau de Elman. Chaque réseau de Elman doit détecter sur sa couche d'output l'invariant qui spécifie la propriété occurrente à laquelle est associé le réseau de Elman. Ensuite, la couche d'output de chacun de ces réseaux de Elman constitue la couche d'input du deuxième niveau. Au deuxième niveau, nous avons un réseau « feedforward » à deux couches ou multi-couches. Nous n'avons pas besoin ici d'un réseau récurrent simple, car les propriétés occurrentes d'une affordance se présentent, non pas successivement, mais simultanément.

Avant de passer au premier chapitre, j'aimerais apporter une réponse à une question importante, celle à savoir si le réseau qui a été décrit ici permet une détection *directe* de l'affordance (l'invariant d'invariant). Un mécanisme qui permet une détection directe est un mécanisme qui n'a pas recours à des inférences. Le réseau de neurones décrit ici s'appuie-t-il donc sur des inférences? Dans ce réseau, aucune inférence n'a lieu. D'une part, sur les

différentes couches, il n'est question que de représentation de structures de perspectives, de représentation d'invariants et de représentation d'invariants d'invariants. D'autre part, les connexions ne représentent pas des règles d'inférences (par ex., des productions, c'est-à-dire des règles ayant la forme « Si A, alors B »). Donc, comme ce réseau ne contient rien d'autre que ces couches et que ces connexions, il ne procède pas sur la base d'inférences. Les invariants d'invariant sont *directement détectés* par ce réseau de neurones.

CHAPITRE VI

LES VARIABLES OPTIQUES ET LE CONTRÔLEUR MOTEUR

Je vais maintenant me pencher sur les notions de variable optique de contrôle et de contrôleur moteur. Je rappelle d'abord ce que font ces deux éléments. En bref, la variable optique de contrôle sert d'input au contrôleur moteur et le contrôleur moteur doit émettre les commandes motrices qui vont activer une synergie de manière à éliminer l'erreur spécifiée par la variable optique. Comment les commandes motrices du contrôleur moteur activent-elles les muscles? Un système agit sur un autre système en affectant les paramètres de contrôle de ce dernier système (Ashby, 1956, 1960) Donc, le contrôleur moteur agit sur les muscles en émettant des commandes motrices qui affectent les paramètres de contrôle des muscles. Donc, pour savoir plus précisément comment les commandes motrices activent les muscles, il faut définir les paramètres de contrôle des muscles. Dans la première partie du présent chapitre, nous allons tenter de définir les paramètres de contrôle des muscles. Dans la deuxième partie, nous allons revenir sur la notion de variable optique de contrôle et nous allons en donner quelques exemples. Enfin, dans la dernière partie, nous allons soulever quelques-uns des problèmes qui touchent le contrôleur moteur et nous allons tenter de les résoudre.

6.1. Les paramètres de contrôle des muscles

Tout d'abord, il faut voir que les commandes motrices n'affectent pas seulement les muscles. Elles affectent aussi l'ossature. Cependant, elles affectent seulement indirectement l'ossature; seuls les muscles sont directement affectés par les commandes motrices. Ce système musculo-squelettique peut être compris comme un système « masse-ressort »; les muscles possèdent les propriétés du ressort (Feldman, 1966; Asatryan et Feldman, 1965;

Latash, 1993). Dans cette première partie, j'aimerais montrer que, lorsque le système musculo-squelettique est compris de la sorte (c.-à-d., compris comme un système « masse-ressort »), les paramètres de contrôle du système sont l'amortissement (damping) et la raideur (stiffness) des muscles.

L'*espace d'état* d'un système est un espace au sein duquel les caractéristiques pertinentes du système sont représentées. Ainsi, si un système peut être représenté par deux types de variables (ou caractéristiques), l'espace d'état de ce système comprendra deux dimensions, une pour chaque type de variable. De manière générale, les systèmes masse-ressort peuvent être représentés dans un espace d'état à deux dimensions. Autrement dit, on peut les décrire selon deux caractéristiques ou variables. Le système masse-ressort est formé d'une substance élastique (un ressort, par exemple) sur laquelle est accrochée une masse. Cette substance élastique peut être allongée ou rapetissée. La première variable qui permet de décrire le système masse-ressort est cet allongement (ou rétrécissement). Suite à l'allongement (ou au rétrécissement), le ressort et la masse oscillent avec une certaine vitesse. La deuxième variable qui nous permet de décrire le système masse-ressort est la vitesse de ces oscillations. L'état de l'oscillateur (le système masse-ressort) correspond donc à un point dans un espace (l'espace d'état) à deux dimensions où les dimensions sont la distance et la vitesse.

Les systèmes dynamiques changent d'état. Les états successifs d'un système forment une *trajectoire* (une succession continue de points) dans l'espace d'état. L'ensemble des trajectoires dans l'espace d'état constitue le *portrait de phase* du système. Le portrait de phase est différent d'un système masse-ressort à l'autre et on peut apprendre à connaître un système masse-ressort en examinant son portrait de phase. Commençons par examiner le portrait de phase de l'oscillateur (le système masse-ressort) le plus simple, *l'oscillateur harmonique.*

Le portrait de phase de l'oscillateur harmonique est composé de cercles concentriques. Cela signifie que si le ressort est allongé d'une certaine distance et que le ressort est ensuite laissé à lui-même, la fin de chaque cycle va correspondre à un allongement qui est identique à l'allongement initial. Cela découle du fait que, dans un oscillateur harmonique classique, il n'y a pas de friction. De plus, les cercles concentriques ont tous la même période. Autrement dit, les cycles correspondant à un cercle concentrique prennent le même temps que les cycles

correspondant à un cercle concentrique d'une grandeur différente. Cela découle du fait que la force produite par le ressort lorsqu'on l'allonge (ou le rapetisse) est linéaire[50].

Il ne faut pas croire que tous les ressorts linéaires ont le même impact sur les oscillations. Un ressort linéaire plus raide (stiffer) qu'un autre ressort linéaire agit différemment sur le comportement du système masse-ressort[51]. Les oscillations d'un premier ressort sont plus rapides que les oscillations d'un second ressort si la raideur du premier ressort est plus prononcée que la raideur du second. Donc, il faut retenir que le paramètre « raideur » contrôle la vitesse des déplacements du ressort.

L'oscillateur harmonique n'est pas un modèle complet pour représenter le système formé par les muscles et l'ossature. Les muscles sont des substances élastiques de type non linéaire (Feldman, 1966) et le déplacement des muscles est freiné dans son déploiement; on y trouve de la friction ou amortissement (Feldman, 1966). Afin de simplifier mon exposé, je n'examinerai pas le cas des ressorts non linéaires. Cependant, pour montrer qu'il est possible de contrôler la destination finale (la position finale) du ressort, je ne peux faire l'économie d'une description des effets de l'amortissement sur les oscillateurs.

Lorsqu'il y a de la friction ou un amortissement, le portrait de phase du ressort n'est plus composé de cercles concentriques, mais de spirales menant toutes vers un même point, le point focal, là où la vitesse est nulle. En variant l'amortissement, on peut changer le point focal. Donc, en variant l'amortissement, on peut faire en sorte que le ressort s'arrête à n'importe quelle distance.

Ainsi, en agissant sur l'amortissement (damping) et la raideur (stiffness) de l'oscillateur on peut décider de l'endroit où il va s'arrêter et à quelle vitesse il va produire son mouvement. De plus, il faut souligner que, pour mener l'oscillateur à destination, il n'est pas requis de surveiller (monitor) son déplacement; l'oscillateur se rend de lui-même à destination. Les paramètres de contrôle des systèmes « masse-ressort » sont donc

---

[50] La linéarité du ressort signifie que la force requise pour allonger (ou rapetisser) le ressort d'une distance X correspond à une constante multipliée par cette distance (F= –kx). Ainsi, si la force requise pour allonger le ressort de 2 mètres est de 10 N, elle sera de 20 N pour l'allonger du double, soit 4 mètres. Dans un graphique représentant la force en fonction de la distance (allongement), la linéarité est représentée par une droite dans ce graphique.

[51] Dans un graphique où la force est fonction de la distance, la raideur du ressort correspond à la pente de la fonction (une droite). Plus la pente est prononcée, plus le ressort est raide.

l'amortissement et la raideur. La raideur détermine la vitesse du ressort et l'amortissement détermine sa position finale, cela sans qu'il soit nécessaire d'effectuer un « monitoring » sur le ressort. Donc, comme le système musculo-squelettique est un système « masse-ressort », le système musculo-squelettique peut être contrôlé en agissant sur l'amortissement et la raideur des muscles. C'est donc sur ces paramètres que les commandes motrices doivent intervenir pour pouvoir contrôler le mouvement. Il reste maintenant à montrer qu'il existe des variables optiques ayant suffisamment d'information pour pouvoir contrôler le mouvement des membres (c.-à-d. des systèmes « structure coordinative-ossature »).

Pour conclure cette première partie, j'aimerais dégager l'un des avantages importants à comprendre le système musculo-squelettique comme un système « masse-ressort ». L'avantage de comprendre le système musculo-squelettique comme un système « masse-ressort » est de pouvoir contrôler ce système sans tenir compte de son état initial (Turvey, 1990). Dans le portrait de phase faisant intervenir de l'amortissement, le point focal est un *attracteur*; tous les états de l'oscillateur conduisent vers ce point focal. Donc, un système de contrôle peut décider de la destination de l'oscillateur sans avoir à se préoccuper des conditions initiales de l'oscillateur; il suffit que le système de contrôle détermine l'amortissement du système « masse-ressort ».

Ce qui vient d'être décrit explique le mouvement discret, c'est-à-dire un mouvement impliquant des membres qui atteignent une destination finale. Le fait de concevoir les membres comme des oscillateurs permet aussi d'expliquer les mouvements répétitifs. Pour que cela soit possible, il suffit de changer l'amortissement (damping) de manière à ce qu'il corresponde à une équation cubique (Abraham et Shaw, 1992). La citation suivante, tirée de Turvey (1990), appuie les conclusions auxquelles nous sommes arrivés.

> [Ballistic] movements can be achieved by setting the muscle linkage [lire : la structure coordinative] to behave like a spring whose equilibrium position is the intended limb position. This springlike organization simplifies control. To set the spring, one needs only to know the final position; knowledge of the initial position is not needed. With proper damping, the limb moves by itself without monitoring and with a speed dependant on the spring constant [c'est-à-dire : la raideur]. By varying the damping and stiffness, movement patterns can be repetitive or discrete. (Turvey, 1990, p. 940)

Il est maintenant temps de se pencher sur les variables optiques de contrôle.

6.2. Les variables optiques de contrôle

Une variable optique de contrôle correspond à un pattern du flux optique. Donc, pour décrire plus précisément ce qui caractérise une variable optique de contrôle, il faut s'attarder sur la notion de flux optique. Le flux optique correspond aux changements de la structure optique qui se produisent lorsque l'agent (l'observateur) se déplace par rapport à l'environnement ou lorsque l'environnement de l'agent se déplace par rapport à l'agent (ou les deux à la fois). On peut représenter ces changements à l'aide d'un champ vectoriel à deux dimensions, où chaque vecteur représente la vitesse à laquelle un élément de la structure optique se déplace par rapport à l'agent (voir figure 6 ci-dessous).

Une variable optique de contrôle, avons-nous dit, est un pattern du flux optique. Comme exemple de variable optique de contrôle, on peut penser au *focus d'expansion*. Lorsque l'agent se déplace, le flux optique prend de l'expansion et le point à partir duquel le flux prend de l'expansion est le focus d'expansion.

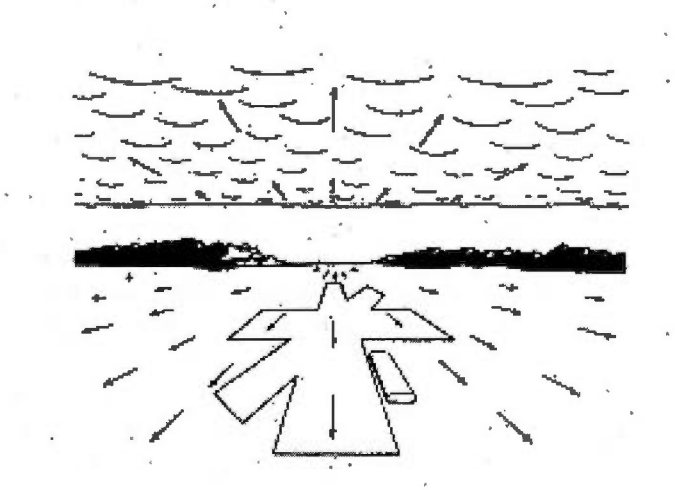


Figure 6.1 Le flux optique lors de la locomotion. Figure tirée de Warren, 1998.

Nous venons de voir qu'une variable optique de contrôle est un pattern du flux optique. Cependant, tout pattern du flux optique n'est pas une variable optique de contrôle. Pour qu'un pattern du flux optique soit une variable optique de contrôle, il suffit et il est nécessaire que ce pattern du flux optique soit corrélé avec le succès (ou l'échec) du mouvement (Fajen, 2007). Plus précisément, à chaque moment du mouvement, la variable optique, selon la

valeur qu'elle prend, doit spécifier (doit donner l'information) si le mouvement se présente comme un succès ou une erreur. Il est clair que si une variable optique satisfait à cette condition, c'est-à-dire si elle permet de dire si le mouvement en cours d'exécution est une erreur ou un succès, cette variable optique permet de contrôler le mouvement. Ainsi, si, lors de l'exécution du mouvement, la variable optique spécifie une erreur, l'agent peut se servir de cette variable pour rectifier son mouvement et, par ce changement moteur, faire en sorte que la variable optique prenne la valeur souhaitée, la valeur qui spécifie un succès. Les variables optiques de contrôle satisfont-elles cette condition? Spécifient-t-elles le succès ou l'échec du mouvement en cours d'exécution? On peut répondre à cette question en procédant à des analyses empiriques. Pour chaque mouvement, on doit chercher dans la structure optique la variable optique qui accompagne le mouvement et voir si cette variable optique spécifie le succès ou l'erreur du mouvement. Les travaux des psychologues écologistes ont en grande partie consisté à trouver de telles variables optiques. J'ai déjà fait allusion à l'une de ces variables optiques : le focus d'expansion du flux optique. Si mon mouvement consiste à se diriger directement sur un objet, le succès de mon mouvement est assuré si je maintiens le focus d'expansion du flux optique au centre de cet objet; et un échec est assuré si le focus d'expansion n'est pas centré sur cet objet. Le focus d'expansion du flux optique est donc une variable optique qui permet de contrôler le mouvement.

Soulignons au passage que cette condition imposée à la variable optique, celle voulant que la variable optique doive spécifier le succès ou l'échec du mouvement, suppose que la variable optique est fonction des mouvements de l'agent. Si la variable optique n'est pas fonction des mouvements de l'agent, la variable optique pourra garder la même valeur, que l'agent exécute un mauvais mouvement (en raison d'une perturbation, par exemple) ou qu'il exécute un bon mouvement; or, si la variable optique garde la même valeur dans un cas comme dans l'autre, elle ne peut pas spécifier le succès ou l'échec du mouvement.

Il existe plusieurs variables optiques de contrôle et celles-ci portent sur des types d'action très variés. Ainsi, on retrouve des variables optiques de contrôle pour l'action de se diriger vers un but (Fajen, 2001; Kim et Turvey, 1999; Wann et Land, 2000), pour l'action de freiner (Lee, 1976; Yilma et Warren, 1995), pour l'action d'attraper une balle au vol (Chapman,

1968; McBeath, Shaffer et Kaiser, 1995) et pour l'action d'intercepter un objet en déplacement (Fajen et Warren, 2004; Peper, Bootsma, Mestre et Baker, 1994).

Deux des variables optiques de contrôle sont celles de Chapman (1968) et celle de Lee (1976). Je vais maintenant présenter brièvement ces deux variables optiques de contrôle.

La variable optique de Chapman se rapporte à l'action de se diriger vers une balle et de l'attraper au vol. Cette variable optique est l'accélération de l'angle d'élévation (α), c'est-à-dire, plus précisément, la dérivée seconde de tan α (c.-à-d. : $d^2 \tan \alpha/dt^2$). L'angle d'élévation (α) correspond à l'angle formé par la balle et l'horizon. Maintenant, quelle valeur doit posséder l'accélération de l'angle d'élévation pour que l'action soit un succès, c'est-à-dire pour que l'agent attrape la balle au vol? Cette accélération doit être nulle. Si l'accélération de l'angle d'élévation est nulle, l'agent est assuré de recevoir la balle à portée de main. Si cette accélération est négative, la balle va tomber devant l'agent; si elle est positive, derrière l'agent.

La variable optique de Lee concerne l'action de freiner. Cette variable optique est la dérivée première de τ (tau), c'est-à-dire le taux de changement de τ. τ est égal au ratio de l'angle optique de l'objet[52] qui approche sur le taux d'expansion de cet angle. La valeur que la dérivée première de τ doit posséder pour que l'action soit un succès, c'est-à-dire pour que l'agent freine en s'arrêtant immédiatement à proximité de l'objet, est de –0,5. Si la dérivée première de τ est inférieure à -0,5, l'agent va percuter l'objet, et si la dérivée première est supérieure à -0,5, il va s'arrêter trop tôt.

Avant d'aller plus loin, j'aimerais revenir sur une proposition qui a été défendue et qui mérite d'être considérée avec plus d'attention. J'ai dit que les variables optiques dont il vient d'être question ne spécifiaient pas des affordances. Les derniers développements nous permettent maintenant de mieux comprendre pourquoi ce n'est pas le cas. Prenons la variable optique de Lee. Celle-ci ne dit pas « il est possible de freiner ». C'est l'affordance qui nous dit cela. En fait cette variable optique nous dit : « si tu continues à agir de la sorte, tu t'arrêteras (ou non) à la distance requise ». Donc, clairement, il faut distinguer entre ces

[52] L'angle optique correspond à l'angle formé par la projection optique des bords de l'objet sur l'œil de l'observateur.

variables optiques et les variables optiques qui spécifient des affordances. La citation suivante de Fajen (2007) confirme le bien-fondé de notre distinction :

> In Lee's (1976) model of braking, tau-dot [la dérivée première de τ] specifies the sufficiency of the current rate of deceleration – that is, it tells the actors "I am (not) braking enough to avoid collision." Likewise, the relevant optical variable in Chapman's (1968) OAC model of fly ball catching (variable) tells the fielder "I am (not) running fast enough to reach the landing location in time." The properties are not affordances – that is, they are not possibilities for action. The affordance for braking is "It is (not) still possible to stop within the limits of my braking capabilities." The affordance for fly ball catching is "It is (not) still possible to reach the landing location on time within the limits of my running capabilities." These are entirely different properties that are specified by entirely different optical variables. (Fajen, 2007, p. 390)

La variable optique, comme nous l'avons vu, sert d'input au contrôleur moteur. Nous allons maintenant nous pencher sur ce dernier.

6.3. Le contrôleur moteur

Dans cette section, je vais donner un avant-goût de l'apprentissage moteur, dont il sera amplement question dans le prochain chapitre, et je vais aborder certains problèmes inhérents au contrôleur moteur.

Il est possible qu'un agent perçoive la bonne variable optique de contrôle et que, malgré cela, il ne parvienne pas à exécuter le bon mouvement, le mouvement qui satisfait la perception de l'affordance. Comment expliquer cela? Pour exécuter le bon mouvement, il faut percevoir la bonne variable optique de contrôle *et* il faut que, après avoir reçu comme input la variable optique de contrôle, le contrôleur moteur émette les commandes motrices qui vont éliminer l'erreur spécifiée par la variable optique de contrôle. Autrement dit, il faut que le contrôleur moteur émette des commandes motrices qui vont activer une synergie de manière à ce que le mouvement exécuté ramène la variable optique à la valeur souhaitée, c'est-à-dire la valeur qui spécifie le succès. Qu'est-ce qui fait que, initialement, le contrôleur moteur ne soit pas en mesure d'émettre ces commandes motrices et que, à la suite de l'apprentissage, il le puisse? C'est évidemment parce que la pondération du réseau de neurones (le contrôleur moteur) a changé en cours de route. Donc, l'apprentissage moteur

doit permettre cela : pondérer le réseau de neurones (le contrôleur moteur) de manière à ce qu'il émette les commandes motrices qui vont activer une synergie de manière à produire un mouvement qui va éliminer l'erreur spécifiée par la variable optique (autrement dit, de manière à donner à la variable optique la valeur souhaitée). C'est au chapitre 7 que nous allons nous pencher sur ce mécanisme d'apprentissage.

Avant de terminer, j'aimerais souligner certains problèmes qui ont été traditionnellement attribués au contrôleur moteur.

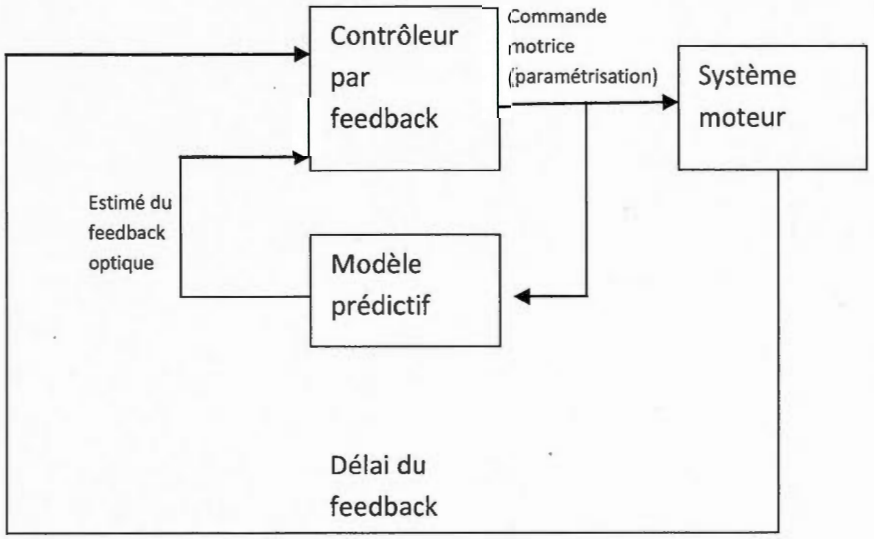


Figure 6.2 Les boucles de feedback interne et externe.

Le modèle du contrôle moteur ayant a été décrit est un modèle du contrôle par feedback. La synergie ou structure coordinative, une fois activée par les commandes motrices, exécute un mouvement. Ce mouvement a des conséquences sur la variable optique. La variable optique, qui a été modifiée par le mouvement, est alors retournée au contrôleur moteur, qui réagira à cette variable optique. Ce *feedback* (la variable optique qui est retournée au contrôleur) se présente comme une erreur (ou un succès). La fonction du contrôleur consiste ensuite à émettre des commandes motrices qui vont activer une synergie dont le mouvement va éliminer l'erreur spécifiée par la variable optique de contrôle. Ce réseau de neurones (le contrôleur moteur) est donc un contrôleur par feedback. Or, le contrôle par feedback est

problématique. Wolpert et ses collègues (Miall et Wolpert, 1996, Miall *et al.*, 1993) soulignent que le feedback est plutôt lent. Pour cette raison, nos mouvements rapides, lorsqu'ils sont produits par ce type de contrôleur, souffrent d'instabilité. Cette objection à notre modèle du contrôle moteur ne signifie pas qu'il faille l'abandonner. Wolpert et ses collègues (Miall et Wolpert, 1996; Miall *et al.*, 1993) ont montré qu'en incorporant un *modèle prédictif* au contrôleur par feedback, l'objection qui vient d'être présentée s'évanouissait.

Les modèles prédictifs prédisent soit le prochain état du système moteur, soit le feedback proprioceptif, soit le feedback optique, c'est-à-dire la valeur de la variable optique après que l'agent ait exécuté son mouvement. Les modèles prédictifs peuvent résoudre le problème qui touche le contrôle par feedback, car l'output (un estimé du feedback optique) des modèles prédictifs est très rapide, beaucoup plus rapide que le feedback optique réel. Donc, pour résoudre le problème susmentionné, il suffit d'instaurer un feedback *interne*, qui met en relation le contrôleur par feedback et le modèle prédictif (voir la figure 6.2). Comme les prédictions du modèle prédictif sont très rapides, le contrôleur peut exécuter des mouvements très rapides en se fiant à ce feedback optique interne (les estimés du feedback optique) plutôt qu'au feedback optique externe. L'insertion d'une boucle de feedback interne contient cependant sont lot de problèmes. Le feedback optique réel, comme nous l'avons dit, souffre d'un délai considérable avant de parvenir au contrôleur. Pour pallier à ce délai, le contrôleur se fie aux estimés du feedback interne (les prédictions du modèle prédictif). Mais qu'arrive-t-il lorsque le feedback *réel* parvient au contrôleur? Cette information est dépassée; sur la base du feedback interne, le contrôleur par feedback a déjà traité cette information. Donc, si le contrôleur par feedback tient compte du feedback optique réel, le contrôleur va émettre une commande motrice erronée. Pour résoudre ce problème, il faudrait pouvoir annuler le feedback optique réel et se fier seulement au feedback optique interne. Mais d'un autre côté, on ne peut pas faire complètement abstraction du feedback optique réel. Le système moteur ou l'environnement peuvent avoir à l'occasion des comportements inhabituels. Lorsque cela se produit, le modèle prédictif va émettre des estimés du feedback optique qui ne correspondent pas à la réalité. Ainsi, lorsque le contrôleur sera informé par ces estimés erronés, celui-ci va émettre des commandes motrices erronées. Dans ces circonstances, il faut que le contrôleur soit informé par le feedback optique réel et qu'il puisse corriger l'action

erronée qu'il aura produite. Le problème consiste donc ici en la coexistence de deux *desiderata* difficilement conciliables. D'une part, il faut pouvoir annuler le feedback optique réel lorsque les prédictions du modèle prédictifs sont justes et, d'autre part, il faut tenir compte du feedback optique réel lorsque les prédictions du modèle prédictif sont erronées, en raison par exemple d'un comportement inattendu du système moteur ou de l'environnement. Wolpert et ses collègues concilient ces deux *desiderata* en incorporant le système de contrôle au sein d'un prédicteur de Smith (Miall *et al.*, 1993; Miall et Wolpert, 1996) (voir figure 6.3).

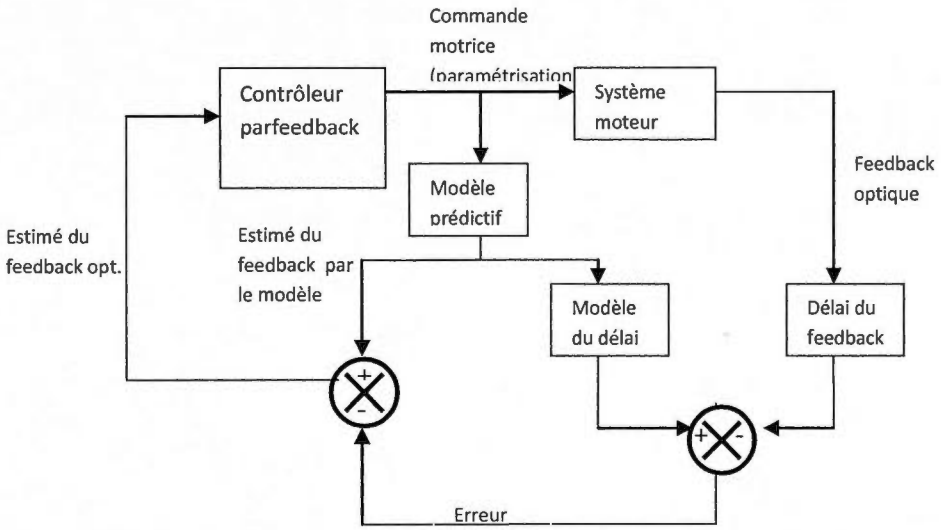


Figure 6.3 Le prédicteur de Smith. Inspiré de Miall et Wolpert, 1996.

Je vais maintenant décrire le fonctionnement du prédicteur de Smith, et, comme nous allons le voir, celui-ci permet de concilier les deux *desiderata* dont il vient d'être question.

Le prédicteur de Smith fait intervenir un modèle du délai. Le modèle du délai est un système qui reçoit comme input l'estimé du modèle prédictif et qui sort comme output ce même estimé, mais quelques instants plus tard. Le temps que prend le modèle du délai pour émettre son output (l'estimé du modèle prédictif) correspond au délai du feedback optique réel. Ensuite, l'output du modèle du délai (l'estimé du feedback optique) ainsi que le

feedback optique réel sont acheminés vers un comparateur. Comme le temps que prend le modèle du délai pour émettre l'estimé du feedback optique correspond au délai du feedback optique réel, l'estimé du feedback optique et le feedback optique arriveront en même temps au comparateur. Donc, si l'estimé du feedback optique est identique au feedback optique réel ou, autrement dit, si l'estimé du feedback optique est juste, cet estimé et le feedback optique réel vont s'annuler au sein de ce comparateur. Dans ce cas de figure, c'est-à-dire lorsque la prédiction du modèle prédictif (l'estimé du feedback optique réel) est valable, le contrôleur va seulement recevoir comme input l'estimé du feedback optique; il ne sera pas perturbé par le feedback optique réel. Le premier *desideratum* énoncé ci-dessus est donc satisfait. Nous souhaitions aussi que si le feedback optique réel est différent de l'estimé du feedback optique, le contrôleur soit touché d'une manière ou d'une autre par ce feedback optique réel (le 2ème desideratum). Le prédicteur de Smith rend précisément compte de cette possibilité. Dans ce cas de figure, c'est-à-dire lorsque le feedback optique réel et l'estimé du feedback optique sont différents, le comparateur est soumis à deux contenus informationnels différents. Le comparateur émettra donc une erreur, qui correspond à la différence entre l'estimé du feedback optique et le feedback optique réel. Cette erreur est ensuite acheminée vers un autre comparateur. Cette erreur parvient au comparateur en même temps qu'un autre estimé du feedback optique. Ce dernier estimé du feedback est un estimé du feedback optique *actuel*. Il survient quelques instants après l'estimé du feedback ayant servi d'input au modèle du délai. Cependant, cet autre estimé, l'estimé du feedback optique actuel, est dépendant de l'estimé antérieur, celui qui a été envoyé au modèle du délai. Donc, l'erreur qui a été établie entre l'estimé antérieur et le feedback optique réel concerne aussi l'estimé ultérieur (l'estimé du feedback optique actuel). En effet, on peut penser que l'erreur qui touchait l'estimé du feedback optique antérieur se soit répercutée sur l'estimé du feedback optique actuel. Donc, en comparant l'estimé du feedback optique actuel et l'erreur, il est possible de trouver un estimé du feedback optique actuel qui soit fiable, c'est-à-dire un estimé du feedback optique qui soit conforme à la réalité. Le comparateur émet donc un estimé du feedback optique actuel qui est fiable (conforme à la réalité). Sur la base de cet estimé du feedback optique, le contrôleur peut corriger son action.

Ce qui précède nous montre comment le système de contrôle corrige son action à partir de la distorsion entre la prédiction du modèle prédictif et le feedback optique réel. Mais il n'est

pas impossible que la correction du mouvement se produise sans qu'on ait besoin de faire intervenir le feedback optique réel. Ce dernier type de correction représente un cas de correction « on-line ». Pour rendre compte de la correction « on-line », il faut, entre autres, faire intervenir un planificateur moteur (Wolpert, 1997; Wolpert et Ghahramani, 2000). Le planificateur moteur reçoit comme input une intention, qui est une représentation *conceptuelle* du mouvement. La tâche du planificateur est de donner une forme *déterminée* à cette intention. Plus précisément, le planificateur doit trouver une « trajectoire » déterminée au mouvement représenté de manière seulement conceptuelle par l'intention. Cette trajectoire correspond à l'état désiré. Cette « trajectoire » ou état désiré est l'input du contrôleur moteur. Le contrôleur moteur reçoit aussi comme input l'*état* actuel du système moteur (par exemple, sa position et sa vitesse). Ici, le contrôleur moteur n'est pas compris comme un contrôleur par feedback, mais comme un modèle inverse. La fonction de ce contrôleur moteur (un modèle inverse) est d'émettre une commande motrice pour chacun des points de la « trajectoire » qui a été calculée par le planificateur moteur. Maintenant, comment rendre compte de la correction « on-line à partir de ces nouveaux éléments »? Encore une fois, la prédiction du modèle prédictif intervient ici. La correction « on-line » se produit à partir de la distorsion entre l'état désiré (la « trajectoire ») et l'état prédit par le modèle prédictif. Ici, il faut souligner que le modèle prédictif ne prédit pas la prochaine valeur de la variable optique mais le prochain *état* du système moteur. Lorsque le modèle inverse est informé de cette prédiction (c.-à-d. informé de l'estimé du prochain état) et qu'il constate que cette prédiction diverge de l'état désiré, il corrige « on-line » le mouvement, c'est-à-dire qu'il corrige le mouvement sans entrer en contact avec le feedback optique réel.

Le modèle du contrôle moteur qui a été présenté dans les dernières pages ne fait pas intervenir d'état désiré (un planificateur moteur) ni de modèle inverse. À ce titre, mon modèle du contrôle moteur est-il en mesure d'apporter une correction « on-line » à l'action, c'est-à-dire une correction qui ne s'appuie pas sur le feedback réel? Absolument. Le contrôleur moteur, dans mon modèle, reçoit comme input l'erreur spécifiée par la variable optique de contrôle. Le contrôleur peut donc corriger son action à partir de cette information sans attendre quoique ce soit du feedback réel.

6.4 Résumé

Dans ce chapitre, nous avons d'abord cherché à savoir comment les commandes motrices activaient les muscles. Plus précisément, comme un système agit sur un autre en affectant ses paramètres de contrôle, nous avons cherché à savoir quels étaient les paramètres de contrôle des muscles. Comme ceux-ci sont analogues à des ressorts, nous avons vu que l'amortissement (damping) et la raideur (stiffness) étaient les paramètres de contrôle des muscles. C'est donc en modulant l'amortissement et la raideur des muscles que les commandes motrices produisent une synergie.

Ensuite, nous avons défini les variables optiques de contrôle. Nous avons vu que celles-ci étaient des patterns du flux optique qui spécifiaient l'erreur ou le succès de l'action. Nous avons ensuite donné deux exemples de variables optiques de contrôle, celle de Chapman et celle de Lee.

Enfin, en troisième lieu, nous avons traité du contrôleur moteur. Nous avons que l'apprentissage moteur, qui concerne le contrôleur moteur, consistait à donner une pondération appropriée au réseau de neurones (le contrôleur moteur). Et, par « pondération appropriée », il faut comprendre une pondération qui génère des commandes motrices qui activent une synergie dont le mouvement permet d'éliminer l'erreur spécifiée par la variable optique de contrôle. Aussi, en dernier lieu, nous avons mis en évidence certains des problèmes qui touchent le contrôleur moteur. Ces problèmes provenaient du fait que le contrôleur moteur était un contrôleur par feedback. Aidés par Wolpert et ses collègues, ces problèmes ont pu être résolus.

2ÈME PARTIE :

LES MÉCANISMES D'APPRENTISSAGE

Dans la première partie de ma thèse, j'ai pris le temps de définir les éléments qui sont communs à l'action de l'expert et du novice. Nous avons d'abord constaté que l'action présuppose la présence d'intentions au sein de l'agent. C'est pour cette raison que, au cours du chapitre 2, j'ai défini le mécanisme qui est responsable de la production des intentions (ou du plan). Aussi, nous avons posé que les deux autres éléments essentiels de l'action, que ce soit celle de l'expert ou du novice, étaient les perceptions d'affordances et les perceptions de variables optiques de contrôle. L'expert et le novice, à travers leurs plans respectifs, contrôlent leurs perceptions d'affordances; autrement dit, leurs plans déterminent (conjointement avec l'environnement) l'affordance à laquelle ils feront attention. Tandis que, chez le novice, ce contrôle s'effectue par l'intermédiaire d'un plan très détaillé (c.-à-d. des intentions très particulières), ce contrôle s'effectue chez l'expert par une ou quelques intentions très générales. Et, quant à la perception de la variable optique, nous avons vu au chapitre 1 qu'elle était contrôlée par la perception de l'affordance; autrement dit, c'est la perception de l'affordance qui détermine (conjointement avec l'environnement) quelle variable optique sera perçue. C'est lors des chapitres 3 et 4 que j'ai défini et défendu la notion d'affordance et de perception d'affordance. Dans le chapitre 5, j'ai tenté de rendre encore plus plausible l'idée qu'il était possible de percevoir des affordances; m'inspirant du modèle récurrent de Elman, j'ai développé un mécanisme qui peut rendre compte de la perception des affordances. Et, dans le chapitre 6, je me suis penché sur la notion de variable optique de contrôle. Comme je viens de le dire, tous les éléments dont il fut question au cours de la première partie de ma thèse — en fait, il fut essentiellement question d'intention, de perceptions d'affordances et de variables optiques de contrôle — se retrouvent autant chez le

novice que chez expert. Dans la seconde partie de ma thèse, mon attention va porter, non plus sur ce qui est commun à l'action experte et à l'action du novice, mais sur les changements qui caractérisent le passage du novice à l'expert. Mon objectif, dans la deuxième partie de ma thèse, est d'expliquer ces changements. Ces changements se produisent à travers des processus d'apprentissage. Donc, mon objectif consiste plus précisément à découvrir les mécanismes d'apprentissage qui sont à la base des différents changements qui caractérisent le passage à l'expertise.

C'est au cours du chapitre 1 que les différents changements qui caractérisent le passage à l'expertise ont été définis. Comme les propos du premier chapitre ont été tenus il y a déjà un certain temps, il serait bon d'en rappeler les grandes lignes.

Lors du premier chapitre, j'ai posé que le passage à l'expertise se caractérisait par quatre grands changements.

Le premier changement est un changement moteur. Le mouvement consiste en une réaction (médiate) des muscles à la variable optique de contrôle. Or, même lorsque la variable optique est appropriée, l'agent peut ne pas exécuter le bon mouvement, celui qui correspond à l'affordance. En effet, si la (bonne) variable optique de contrôle est l'input d'un contrôleur moteur dont la pondération est mauvaise, alors le mouvement de l'agent sera inapproprié quand bien même il perçoit la bonne variable optique. Le premier changement correspond donc à un changement du poids des connexions du réseau de neurones (le contrôleur). Lors de ce changement, le réseau de neurones passe d'une pondération qui génère des commandes motrices dont le mouvement ne permet pas d'éliminer l'erreur spécifiée par la variable optique à une pondération qui génère des commandes motrices dont le mouvement permet d'éliminer l'erreur spécifiée par la variable optique de contrôle. J'ai trouvé chez Ashby, dans son texte *Design for a Brain* (1960), un mécanisme d'apprentissage qui peut rendre compte de ce changement de pondération. C'est au cours du chapitre 7 que je vais présenter ce mécanisme d'apprentissage.

Le deuxième changement est un changement perceptuel. Au tout début, l'agent ne perçoit pas les bonnes affordances ni les bonnes variables optiques de contrôle[53]. C'est ainsi que, lorsqu'il commence à apprendre son sport, le frappeur au baseball va percevoir la balle qui se dirige vers lui comme un objet qui « afforde » le danger ou le choc, tandis que, lorsqu'il est expert, cette même balle va se présenter comme quelque chose qui « afforde » l'action d'être frappé. Le même genre de constat vaut pour la variable optique de contrôle. Ainsi, le voltigeur novice ne perçoit pas l'accélération de l'angle d'élévation de la balle (la variable optique de contrôle), tandis que, une fois expert, il perçoit cette variable optique de contrôle. Un second mécanisme d'apprentissage doit donc être posé ici. Celui-ci doit nous expliquer comment on passe de la perception d'une affordance (ou d'une variable optique) inappropriée à la perception d'une affordance (ou d'une variable optique) appropriée. J'ai trouvé que l'apprentissage par renforcement tel que présenté par Sutton et Barto (1998) permettait d'expliquer ce changement perceptuel. C'est dans le chapitre 8 que je vais présenter ce mécanisme d'apprentissage.

Le troisième changement d'importance qui caractérise le passage à l'expertise concerne la diminution progressive du plan. Les perceptions d'affordances, comme on l'a vu au chapitre 1, sont contrôlées par des intentions (un plan). Lorsque l'agent est novice, il a besoin d'un plan détaillé (un nombre considérable d'intentions très particulières, c'est-à-dire des intentions qui se situent juste au-dessus du niveau des habitudes et des habiletés) pour contrôler les perceptions d'affordances; et plus il devient expert, moins le contrôle des perceptions requiert un plan détaillé; à la limite, une seule intention suffit pour contrôler les perceptions initialement contrôlées par une multitude d'intentions particulières. Ainsi, l'expert au tennis pourrait contrôler toutes ses perceptions d'affordances à l'aide d'une seule intention, disons l'intention de gagner le point en cours. C'est grâce à un processus d'apprentissage que l'agent devient réactif, c'est-à-dire devient réactif à des perceptions d'affordances qui sont sous le contrôle d'un plan réduit à sa plus simple expression. En fait, c'est le système perceptuel qui doit passer par un processus d'apprentissage. Plus précisément, le système perceptuel doit apprendre à émettre les mêmes outputs (les mêmes

[53] Les affordances sont bonnes ou mauvaises selon qu'elles permettent ou non de satisfaire l'intention. Ainsi, si l'affordance permet peu ou pas de satisfaire l'intention, nous dirons qu'il s'agit d'une mauvaise affordance et, inversement, qu'il s'agit d'une bonne affordance si elle satisfait l'intention.

perceptions d'affordances) à partir d'un nombre beaucoup plus restreint d'inputs (les intentions générales). L'apprentissage qui est à la base du troisième changement est donc un apprentissage perceptuel. Or, selon ce qui a été dit plus tôt, le type d'apprentissage qui se produit au sein du système perceptuel est un apprentissage par renforcement. Donc, il y a tout lieu de penser que l'apprentissage qui est à la base du troisième changement est un apprentissage par renforcement. C'est au cours du chapitre 9 que je vais montrer que l'apprentissage par renforcement permet d'expliquer que l'agent passe d'une action qui est sous le contrôle d'un plan détaillé à une action qui est sous le contrôle d'un plan beaucoup plus dépouillé.

Le quatrième type de changement concerne l'amélioration du plan[54].

Lorsqu'il est novice, l'agent fait reposer son action sur un plan détaillé. Au tout début, le novice élabore des mauvais plans et, plus il apprend, plus ses plans sont efficaces. L'apprentissage qui est à la base de l'amélioration du plan ne fera pas l'objet d'un chapitre en particulier. Je vais me contenter d'esquisser les caractéristiques générales de cet apprentissage dans ce qui suit.

À la lumière de ce qui a été dit à propos de la production des intentions ou des plans, on peut d'emblée tirer certaines conclusions à propos de cet apprentissage. À la lumière du chapitre 2, nous savons que les intentions (ou plans) sont produites par le GPS. Donc, pour que l'agent ait de meilleurs plans, il faut que le GPS produise de meilleurs plans. Pour que le GPS produise de meilleurs plans, il faut modifier certaines de ses règles ou productions. Donc, l'apprentissage qui est à la base de l'amélioration du plan doit nous montrer comment le GPS modifie certaines de ses règles. La modification de ces règles peut provenir d'un apprentissage supervisé. L'instructeur suggère à l'agent les règles qu'il doit changer et l'agent se conforme à ces suggestions en changeant les règles en question. Le changement de ces règles peut également faire l'objet d'un apprentissage non supervisé. Par lui-même, c'est-à-dire en raisonnant, l'agent peut arriver à la conclusion que de nouvelles règles sont préférables et il peut remplacer ses anciennes règles par ces nouvelles règles. Le changement des règles peut également provenir d'un apprentissage par renforcement. En procédant de

[54] En fait, comme je l'ai montré au chapitre 1 lorsque j'ai décrit l'ordre d'apparition des changements, le changement qui consiste en une amélioration du plan précède le changement qui consiste en une élimination progressive du plan.

manière exploratoire, l'agent adopte certaines règles et il examine les récompenses qui découlent de l'adoption de ces règles. Il retient ensuite les règles qui rapportent un maximum de récompenses. Il existe probablement d'autres processus d'apprentissage qui permettent l'insertion de nouvelles règles au sein du GPS, mais, dans cette thèse, j'ai décidé de ne pas m'y attarder.

CHAPITRE VII

L'APPRENTISSAGE MOTEUR

Lorsqu'il est novice, l'agent base son action sur un plan très détaillé. L'intention qui est au bas de la hiérarchie cause une perception d'affordance, et cette dernière perception cause la perception d'une variable optique. Comme l'agent est un parfait novice, cette variable optique peut être inadéquate. Autrement dit, il est possible que cette variable optique ne permette pas la réalisation motrice du mouvement auquel réfère l'affordance. Cependant, afin de simplifier notre description de l'apprentissage moteur, nous allons supposer ici que cette variable optique est appropriée. Ainsi, si le voltigeur perçoit l'affordance « attrape-moi » présentée par la balle de baseball frappée au champ centre, cette perception dirigera son attention vers la variable optique appropriée, c'est-à-dire vers la variable optique « accélération de l'angle d'élévation formé par la balle et l'horizon ». À la fin du présent chapitre, je prendrai le temps de décrire ce qui se produit au sein du mécanisme d'apprentissage lorsque la variable optique de contrôle n'est pas appropriée.

La variable optique, même si elle est appropriée, peut autant causer un mauvais mouvement qu'un bon mouvement; tout dépend de la pondération du réseau de neurones (le contrôleur moteur). L'objectif, maintenant, est de trouver un mécanisme d'apprentissage qui pondère le réseau de neurones de manière à ce que celui-ci produisent des commandes motrices qui vont produire un mouvement ayant la particularité d'éliminer l'erreur spécifiée par la variable optique de contrôle ou, autrement dit, la particularité de donner à la variable optique de contrôle la valeur souhaitée. C'est ce type de changement paramétrique (ce changement du poids des connexions) que notre mécanisme d'apprentissage doit permettre d'expliquer. Pour trouver ce mécanisme d'apprentissage, je me suis inspiré du mécanisme d'apprentissage que Ashby présente dans son texte *Design for a Brain* (1960).

J'ai divisé ce chapitre en deux parties. Dans la première partie, je présente quelques notions de base de la théorie des systèmes dynamiques. Ce détour par la théorie des systèmes dynamiques est nécessaire dans la mesure où le mécanisme d'apprentissage de Ashby est en grande partie dynamique. Après avoir exposé les bases de la théorie des systèmes dynamiques, je définis mon mécanisme d'apprentissage, qui, je le répète, est très proche de celui de Ashby (1960).

7.1. Aperçu sur la théorie des systèmes dynamiques[55]

Le concept fondamental de la théorie des systèmes dynamiques est celui de changement. Habituellement, les différentes expositions de la théorie des systèmes dynamiques sont concernées par des changements continus, c'est-à-dire des changements qui se produisent par sauts infinitésimaux, tant temporellement que sur le plan des différences de contenu. Selon Ashby (1956), les sauts infinitésimaux posent trop de problèmes d'ordre mathématique pour une *introduction* à la théorie des systèmes dynamiques. Dans *An Introduction to Cybernetics,* les changements qui intéressent Ashby sont des changements discontinus, c'est-à-dire des changements qui se produisent par des sauts finis, et non pas infinitésimaux.

Selon Ashby, il est préférable d'introduire la théorie des systèmes dynamiques en tenant compte de changements discontinus, d'une part, parce que toutes les questions importantes peuvent être tranchées par un simple comptage, ce qui n'est pas aussi simple lorsqu'il est question de changements continus, puisqu'il faut alors comparer des quantités infinitésimales ou additionner ensemble un nombre infini de quantités infinitésimales; et, d'autre part, parce que l'on ne perd rien en considérant des différences finies. En effet, même les différences continues peuvent être appréhendées à partir de différences finies. La stratégie pour appréhender les différences continues ou infinitésimales à partir de différences finies est fort simple. On s'assure d'abord de savoir ce qui se produit lorsque les différences possèdent une certaine grandeur. On établit ensuite ce qui se passe pour des différences plus petites. Et on répète ensuite cette démarche pour des différences encore plus petites. En continuant de la

---

[55] Ma présentation de la théorie des systèmes dynamiques est inspirée du texte de Ashby, *An Introduction to Cybernetics* (1956).

sorte, il est possible d'appréhender ce qui se passe lorsque les différences tendent vers zéro. Des différences infinitésimales sont alors appréhendées.

Maintenant que j'ai souligné que les changements seront modélisés, non pas de manière continue, mais de manière discontinue, je peux annoncer les actions discursives que je compte réaliser au cours de mon exposition de la théorie des systèmes dynamiques. Premièrement, je définirai le comportement des systèmes dynamiques. Deuxièmement, je définirai l'allure générale des mécanismes qui sont à la base du comportement des systèmes dynamiques.

7.1.1. Le comportement des systèmes dynamiques

Supposons un système quelconque. À chaque instant, celui-ci se trouve dans un état donné. Ce système peut changer d'état. Suivant la terminologie de Ashby, je vais nommer « transition d'état » le passage d'un état à un autre. Le premier état d'une transition d'état est nommé par Ashby « opérant » (operand), tandis que le second état est nommé « transformé » (transform). Le comportement du système se ramène à l'*ensemble* de ses transitions d'état.

Le comportement des systèmes dynamiques peut être représenté de plusieurs façons. Si le comportement est continu, une équation différentielle peut faire l'affaire; sinon (si le comportement est discontinu), la représentation peut prendre la forme d'une équation algébrique non différentielle. La représentation du comportement peut aussi être graphique. Le comportement est alors représenté par le *flux* d'un *espace d'état*. Le mode représentationnel qui sera privilégié au cours du présent chapitre est qualifié par Ashby de tabulaire. Il serait donc souhaitable de le maîtriser. À cet effet, je vais décrire le comportement fictif d'un système fictif et je vais ensuite représenter ce comportement sous forme tabulaire.

Supposons un système quelconque, disons un animal, et attribuons lui le comportement fictif suivant : quand il est affamé (état 1), l'animal cherche de la nourriture (état 2); du moment que la nourriture est trouvée, il se nourrit (état 3); après s'être nourri, il s'endort (état 4); à son réveil, il est de nouveau affamé — il part donc de nouveau à la recherche de nourriture; etc. Représentons maintenant ce comportement (l'ensemble des transitions d'état du système) sous forme tabulaire.

Si l'on suppose que les symboles A, B, C et D représentent respectivement les états 1, 2, 3 et 4 du comportement décrit au paragraphe précédent, la représentation tabulaire du comportement aurait alors la forme suivante :

A B C D
↓ B C D A

S'ils sont situés au-dessus du trait horizontal, les symboles désignent les états du système avant une transition d'état; s'ils sont situés au-dessous du trait horizontal, l'état du système à la suite d'une transition d'état. Ainsi, ce tableau nous dit que si le système est en B, il va passer en C (une première transition d'état); et que, à partir de C, il va ensuite passer en D (une deuxième transition d'état); et que, de D, il va ensuite passer en A (une troisième transition d'état), etc.

Maintenant que la notion de comportement a été esquissée et que nous savons comment représenter le comportement d'un système sous forme tabulaire, nous pouvons passer à la prochaine étape, soit caractériser le comportement des systèmes dynamiques.

Le comportement des systèmes dynamiques est un comportement à états déterminés (state-determined behavior) ou, autrement dit, un comportement de type déterministe (Ashby, 1956; Giunti, 1995). Il faut donc identifier ce qui caractérise un comportement déterministe.

Le comportement d'un système est déterministe quand chacune de ses transitions d'état est *univoque*. L'univocité des transitions d'état signifie que l'état succédant à un état donné reste toujours le même; qu'il ne peut pas changer d'une fois à l'autre. Ainsi, relativement au comportement de l'animal décrit ci-dessus, nous étions en face d'un comportement déterministe dans la mesure où l'état correspondant à la recherche de nourriture fait invariablement suite à l'état « être affamé » et que l'état « se nourrir » succède invariablement à l'état « chercher de la nourriture », etc. Donc, le comportement d'un système ne serait pas déterministe si on pouvait le représenter à l'aide de la représentation tabulaire suivante :

A B C D
↓ AB C D A

Cette représentation tabulaire signifie effectivement que, si le système est en A, il peut passer ou bien en A ou bien en B, ce qui correspond à une transition non univoque.

Il faut aussi noter ceci à propos du caractère déterministe du comportement : un état donné peut très bien succéder à plus d'un état. Autrement dit, pour qu'il y ait un comportement déterministe, les transitions d'état n'ont pas besoin d'être *bi*univoques ou « one-to-one » : l'univocité suffit. Ainsi, le comportement représenté par le tableau suivant est déterministe, quand bien même un même état succède à plus d'un état (opérant).

A	B	C	D
↓			
A	A	A	B

Dans *Introduction to Cybernetics*, Ashby fait remarquer que le comportement d'un système dynamique n'est pas seulement déterministe; son comportement doit aussi être fermé (closed). Lorsque le comportement d'un système est fermé, les transitions d'état du système conduisent toujours à des transformées (le second état d'une transition d'état) qui sont eux-mêmes des opérants (le premier état d'une transition d'état); toutes les transformées du comportement sont aussi des opérants. Ainsi, la représentation tabulaire du comportement suivant n'est pas la représentation d'un comportement fermé

A	B	C	D
↓			
B	C	D	E

Il n'est pas difficile de voir pourquoi le comportement d'un système dynamique doit être un comportement fermé. Un système dynamique adopte toujours un certain comportement. À la limite, le système reste lui-même, ce qui pourrait être représenté par la transition A → A. Or, dans un comportement ouvert, le système peut tomber dans un état qui n'est l'opérant d'aucune transition, si bien que le système cesse d'adopter un comportement. Dans la représentation tabulaire ci-dessus, c'est ce qui se produit lorsque le système passe à l'état E.

Un comportement déterministe possède une autre caractéristique importante : pour connaître le comportement futur du système, il n'est pas utile de connaître l'histoire du système; l'état *actuel* du système suffit (Van Gelder, 1995). Les représentations tabulaires

présentées ci-dessus répondent à cette condition, puisque, dans ces représentations tabulaires, le comportement futur du système peut être connu à partir de l'opérant *actuel*.

J'aimerais rajouter quelque chose d'important à propos du comportement des systèmes dynamiques. Ceux-ci ne sont pas seulement des systèmes ayant un comportement déterministe; leur comportement déterministe est aussi en fonction du temps. Je rajoute cette condition à ma définition des systèmes dynamiques car, à défaut de cette condition, nous sommes obligés d'inclure les systèmes computationnels au sein des systèmes dynamiques. En effet, les systèmes computationnels ont aussi un comportement déterministe (Van Gelder, 1995); ils répondent à toutes les conditions qui ont été énumérées ci-dessus, soit 1) transitions d'état univoques, 2) comportement fermé et 3) transitions d'état indépendantes du passé. Cependant, les transitions d'états des systèmes computationnels ne sont pas fonction du temps; en effet, que la transition d'état d'un système computationnel prenne un instant ou une journée, cela ne vient pas jouer sur la validité de la transition d'état; la transition d'état est aussi valable dans un cas comme dans l'autre. Donc, afin de distinguer système dynamique et système computationnel, je pose que l'un (le système dynamique) possède un comportement déterministe qui est fonction du temps, tandis que l'autre (le système computationnel) possède un comportement déterministe qui est indépendant du temps.

Donc, tous les systèmes dynamiques possèdent un comportement qui est déterministe, fermé et en fonction du temps. Il existe cependant plusieurs types de comportements déterministes. Certains sont chaotiques, mais la plupart sont stables. Au cours de ce texte, je serai essentiellement concerné par des systèmes dynamiques ayant un comportement stable. À ce titre, je vais conclure cet exposé général sur le comportement des systèmes dynamiques par une discussion portant sur la notion de stabilité.

Pour comprendre la notion de stabilité, la notion d'*équilibre* dynamique doit d'abord être définie.

Il existe plusieurs types d'équilibres dynamiques. Les plus communs sont l'équilibre ponctuel et l'équilibre cyclique. L'équilibre d'un système dynamique est ponctuel quand le système atteint un état et s'y maintient. Ainsi, dans la représentation tabulaire suivante, le système est en équilibre ponctuel du moment qu'il atteint l'état A.

A B C D	
↓	
A C D B	

L'équilibre du système dynamique est cyclique quand le système passe par une séquence d'états et qu'il répète de manière ininterrompue cette séquence. Dans la représentation tabulaire précédente, l'équilibre cyclique est atteint du moment que l'état B ou C ou D est atteint. Par exemple, s'il est dans l'état B, le système répétera indéfiniment le cycle « B – C – D ».

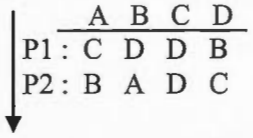
La notion d'équilibre dynamique étant définie, la notion de comportement stable peut maintenant être abordée. Le comportement d'un système est stable lorsque, peu importe l'état initial du système, le système se comporte de manière à atteindre un équilibre dynamique. Inversement, un comportement est instable lorsque, à partir d'un état initial distinct de l'état d'équilibre, le système n'arrive pas à trouver son équilibre. On peut illustrer à l'aide d'un exemple la différence entre comportement stable et comportement instable. Prenons un cône et un cube. Le cube est dans un état d'équilibre du moment qu'il repose sur l'une de ses faces et qu'il y reste; le cône est lui aussi en équilibre s'il repose sur son sommet et s'y maintient. Maintenant, il est clair que, contrairement au cube, le cône n'atteindra pas son état équilibre si son état initial diverge, ne serait-ce que très légèrement, de son état d'équilibre[56]. Le comportement du cône est donc instable; celui du cube, non. Il faut souligner ici qu'il est rare de rencontrer un comportement complètement stable, c'est-à-dire un comportement qui conduit le système vers un état d'équilibre, *peu importe l'état initial*. En fait, dans la plupart des systèmes stables, lorsque l'état initial est très éloigné de l'état d'équilibre, le système n'atteint pas son état d'équilibre. On dit alors que l'état initial du système est en dehors du bassin d'attraction du système. Dans un portrait de phase, le bassin d'attraction correspond à la région qui entoure l'état d'équilibre et à partir de laquelle le système atteint toujours son état d'équilibre (ponctuel ou cyclique).

Avant de conclure cette discussion sur la notion de comportement, il serait important de dire quelques mots sur la notion de paramètre, cette dernière étant impliquée dans la notion de comportement.

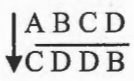
---

[56] En fait, le cône va trouver un autre état d'équilibre. Ce n'est cependant pas cet équilibre qui nous intéresse ici, mais celui qui correspond à l'état du cône lorsque ce dernier se maintient sur la pointe de son sommet.

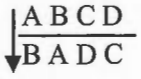
Le comportement d'un système dynamique dépend des conditions (internes ou externes) dans lesquelles le système évolue. Si ces conditions changent d'état, le comportement du système change aussi. Les conditions dans lesquelles le système évolue sont aussi nommées « paramètres » ou « inputs ». Dans une représentation tabulaire, voici comment la dépendance du comportement à l'endroit des paramètres peut être représentée :



Ce tableau nous dit que si l'état des paramètres correspond à P1, le système se comporte de la manière suivante :



tandis que s'ils correspondent à P2, le système se comporte ainsi :



Maintenant que nous en savons un peu plus sur le comportement des systèmes dynamiques, il est temps de se pencher sur les mécanismes qui sont à la base de leur comportement.

7.1.2. Le mécanisme qui est à la base du comportement d'un système dynamique

Pour déterminer l'allure générale des mécanismes qui sont à la base du comportement des systèmes dynamiques, je vais procéder en deux temps. Dans un premier temps, je vais déterminer les propriétés générales de ces mécanismes. Dans un deuxième temps, je vais donner un contenu plus intuitif aux considérations générales de la première partie : je vais décrire un mécanisme particulier et montrer comment ce mécanisme génère le comportement du système dynamique.

7.1.2.1. Les propriétés du mécanisme des systèmes dynamiques

Expliquer mécaniquement le comportement d'un système ou, autrement dit, trouver le mécanisme responsable du comportement d'un système, c'est ni plus ni moins montrer que le comportement en question découle de la façon dont les parties du système agissent les unes sur les autres (Zednik, 2011). Dans un système dynamique, les parties du système interagissent entre elles (Van Gelder, 1995). Dire que des parties interagissent, c'est dire que l'action de chaque partie affecte (immédiatement ou médiatement) toutes les autres. Cette première caractérisation du mécanisme des systèmes dynamiques, c'est-à-dire le fait que ce mécanisme suppose une interaction entre les parties du système, n'est pas suffisante. En effet, le mécanisme de n'importe quel système, et pas seulement celui des systèmes dynamiques, suppose une interaction entre les parties du système. Autrement dit, par définition, un système, qu'il soit dynamique ou non, est un ensemble d'objets en interaction (Van Gelder, 1995). Notre objectif, dès lors, est de déterminer les propriétés générales qui sont propres ou spécifiques au mécanisme des systèmes dynamiques.

Comme on l'a vu, les systèmes dynamiques se caractérisent par un comportement de type déterministe. Donc, pour déterminer les caractéristiques spécifiques du mécanisme des systèmes dynamiques, il suffit de préciser la nature des interactions entre les parties du système et de montrer que les précisions qui sont apportées à ces interactions permettent de rendre compte d'un comportement déterministe.

L'interaction entre les parties d'un système est en mesure de rendre compte d'un comportement dynamique à deux conditions. La première condition concerne les parties du système ou, puisque les parties du système coïncident avec les parties du mécanisme, du mécanisme; la seconde condition concerne l'interaction proprement dite entre ces parties.

Abordons d'abord la première condition. Les parties d'un système sont elles-mêmes des systèmes (en fait, des sous-systèmes). En effet, chacune des parties d'un système est composée de parties en interaction[57]. Tel qu'annoncé ci-dessus, nous devons maintenant nous demander à quelle condition les parties d'un système dynamique peuvent générer un comportement déterministe chez le système dynamique. Ces parties doivent-elles être elles-

[57] Ceci est faux si les parties en question correspondent aux éléments ultimes de la matière. Je ne tiendrai pas compte de cette possibilité ici.

mêmes des systèmes déterministes? Ou, au contraire, ces parties peuvent-elles avoir un comportement qui n'est pas déterministe?[58] Clairement, si le comportement des parties d'un système dynamique n'est pas déterministe, le comportement du système dynamique total, celui qui est composé des parties en question, ne saurait l'être non plus. Donc, les parties d'un système dynamique possèdent un comportement déterministe. Et comme les parties du système dynamique coïncident avec les parties du mécanisme qui est à la base du comportement du système dynamique, il faut conclure que les parties de ce mécanisme possèdent un comportement déterministe. Voilà pour la première condition.

Je vais maintenant aborder la deuxième condition, qui concerne le type d'interaction qui est requis pour que le comportement du système soit déterministe[59].

Deux parties interagissent lorsque chacune d'elles agit sur l'autre. Pour spécifier la façon dont deux parties interagissent, il suffit donc de spécifier la nature de l'action qui s'exerce de part et d'autre des parties. Comme je m'intéresse à l'interaction qui génère un comportement déterministe, je cherche plus précisément à déterminer le type d'action que chaque partie doit exercer sur l'autre afin que le système total (le système composé de ces parties) puisse adopter un comportement déterministe.

Une partie agit sur une autre partie si et seulement si l'état de la première partie, celle qui est active, détermine le comportement[60] de la seconde, qui est passive[61]. Ainsi, si la partie A(ctive) agit sur une partie P(assive) et que la partie A change d'état, la partie P devrait changer de comportement. L'action d'une partie peut prendre deux formes : elle peut être univoque ou non univoque. L'action est univoque lorsqu'un état donné de la partie active ne cause jamais plus d'un comportement chez la partie passive. Autrement dit, si la partie A agit de manière univoque sur la partie P, la partie P ne peut pas avoir des comportements différents si la partie A garde le même état. Et, à l'inverse, l'action est non univoque lorsqu'un même état de la partie A peut causer des comportements différents chez la partie P.

---

[58] Il est probable que tout système réel soit déterministe; que, dans le monde, il n'y ait pas de système stochastique. Nos *modèles*, par contre, ne sont pas tous déterministes. En effet, certains de nos modèles sont stochastiques. Donc, lorsque je me demande si les parties d'un système déterministe sont aussi des systèmes de type déterministe, je me situe non pas au niveau de la réalité, mais au niveau des modèles.

[59] Pour simplifier mon analyse, je vais me pencher sur un système qui ne comprend que deux parties.

[60] Il faut se rappeler qu'un comportement a été défini comme un ensemble de transitions d'état.

[61] Afin de simplifier ma description, je vais décrire non pas l'action de chaque partie, mais l'action d'une seule partie.

Dans un système dynamique, il est clair que l'action des parties du système doit être univoque. Si cette action n'était pas univoque, les parties du système adopteraient des comportements différents sous l'effet d'une même action. Or, pour des raisons qui sont évidentes, le comportement du système total ne peut être déterministe si le comportement de ses parties varie lorsqu'elles sont soumises à des conditions identiques.

Ces deux conditions permettent d'expliquer que le système total (le système composé des deux parties) ait un comportement déterministe. Mais le comportement d'un système dynamique est davantage qu'un comportement déterministe; c'est aussi un comportement qui est fonction du temps. La question à se poser, dès lors, est celle-ci : pour que le comportement du système total soit fonction du temps, le comportement des parties, ainsi que l'interaction entre ces parties, doivent-ils eux aussi se dérouler en fonction du temps. Il n'est pas difficile de voir que c'est effectivement le cas. Si le comportement des parties ainsi que leurs interactions ne se déployaient pas en fonction du temps, le comportement du système ne pourrait pas se déployer en fonction du temps. Nous pouvons donc ajouter une troisième condition au mécanisme des systèmes dynamiques : le mécanisme (le comportement des parties ainsi que leurs interactions) d'un système dynamique doit se dérouler en fonction du temps.

Maintenant que les conditions générales du mécanisme des systèmes dynamiques ont été définies, il est temps de donner une dimension plus intuitive à ce qui vient d'être dit à propos de ce mécanisme. Je vais d'abord définir un mécanisme particulier répondant aux conditions énoncées à l'instant et je vais ensuite montrer comment un comportement dynamique peut être produit à partir de ce mécanisme. Aussi, afin de ne pas m'encombrer de considérations empiriques qui pourraient compliquer les choses, ma description va s'inscrire dans un cadre formel. Plus précisément, je vais montrer, non pas qu'un comportement ayant tel contenu particulier découle d'un mécanisme ayant tel contenu particulier, mais plutôt que la *forme* particulière de ce comportement découle de la *forme* particulière de ce mécanisme.[62] Aussi, dans la description de ce mécanisme, je ferai abstraction de la dimension temporelle ou, plus précisément, je *présupposerai* que le comportement des parties ainsi que leur interaction sont dépendantes du temps, sans toutefois expliciter ces dépendances par rapport au temps. Donc,

[62] Précisons ceci : le mécanisme et le comportement du système dynamique sont fictifs. La forme du mécanisme et la forme du comportement sont donc tout autant fictives.

le mécanisme particulier que je vais décrire va se contenter d'expliciter les deux premières conditions décrites ci-dessus, celles qui rendent compte du caractère déterministes des systèmes dynamiques. Je le rappelle, les deux premières conditions qui ont été attribuées au mécanisme des systèmes dynamiques sont : 1) que le comportement des parties du mécanisme doit être déterministe; et 2) que l'interaction entre les parties du mécanisme est univoque.

7.1.2.2. Une explication mécanique particulière (et formelle)

Je vais procéder ici en deux étapes : 1) je vais définir la forme d'un mécanisme répondant aux deux premières conditions ou propriété énoncées ci-dessus, et 2) je vais montrer que le comportement déterministe (plus précisément : la forme de ce comportement) d'un système dynamique découle de (la forme de) ce mécanisme.

Si le mécanisme d'un système satisfait les deux conditions énoncées ci-dessus, ce mécanisme consiste en une interaction univoque entre des parties dotées d'un comportement déterministe. Donc, pour définir quant à sa forme le mécanisme d'un système dynamique, il suffit de définir formellement l'interaction univoque des parties du système ainsi que le comportement déterministe de ces parties. Commençons par définir (formellement) le comportement déterministe des parties du système. Pour simplifier notre démonstration, nous allons supposer que ce système possède seulement deux parties. Voici la définition formelle du comportement de ces deux parties[63] :

Partie X :

	A	B	C	D
Y1 :	C	D	D	B
Y2 :	B	A	D	C
Y3 :	D	C	D	B

Partie Y :

	I	J	K
X1 :	I	I	J
X2 :	K	J	I
X3 :	K	K	K
X4 :	J	K	K

[63] Ces deux comportements se lisent comme le comportement présenté à la fin de la section 7.1.1.

Notons au passage que ces deux définitions satisfont la première condition énoncée ci-dessus. Selon cette condition, le comportement des parties d'un système dynamique doit être déterministe, ce que ces deux tableaux (les deux définitions) permettent clairement d'établir, puisque, pour chaque paramètre, les transitions d'état sont univoques (et indépendantes du passé).

Maintenant que le comportement des deux parties du système a été défini, il reste à définir, toujours dans une perspective formelle, l'interaction qui existe entre ces parties.

L'action de X sur Y consiste à provoquer chez Y un comportement déterminé. Mais X ne provoque pas *directement* le comportement de Y. En effet, nous avons vu que ce qui est directement responsable du comportement d'un système, c'est la valeur de ses paramètres. Donc, ce qui cause directement le comportement de Y, ce sont les valeurs paramétriques X1, X2, X3 et X4. Donc, c'est en agissant causalement sur les paramètres du système Y que le système X agit causalement sur le système Y et qu'il provoque un comportement chez Y (Ashby, 1956). On peut donc définir l'action de X sur Y en définissant l'action de X sur les paramètres de Y. Plus précisément, on peut définir l'action de X sur Y en établissant, pour chaque état de X, la valeur qui est causée chez paramètres de Y. Et l'action de Y sur X correspondrait à l'action causale de chacun des états de Y sur les paramètres de X. Formellement, voici comment j'ai défini l'action de X sur Y et l'action de Y sur X.

L'action de X sur Y :	État de la partie X :	A	B	C	D
	Valeur du paramètre de Y :	X1	X2	X3	X4

L'action de Y sur X :	État de la partie Y :	I	J	K
	Valeur du paramètre de X :	Y1	Y2	Y3

Ces deux tableaux, qui définissent l'action exercée par chacune des parties du système, répondent à la deuxième condition énoncée ci-dessus, selon laquelle l'action d'une partie doit être univoque. L'action d'une partie est univoque si aucun de ses états ne provoque plus d'un comportement au sein de l'autre partie. Pour que cela soit le cas, il suffit de s'assurer que chaque état ne cause pas plus d'une valeur paramétrique. Les deux tableaux précédents établissent précisément cela.

Le comportement des parties du système et l'interaction entre ces parties étant définis, il est temps de passer à la prochaine étape : montrer comment l'interaction entre ces parties produit un comportement de type déterministe au sein du système qui est composé de ces parties.

Supposons que le système composé des parties X et Y — nommons-le « système "Z" » — se trouve initialement dans l'état (A, J). Compte tenu de cet état du système total, la partie X agit sur Y en réalisant le paramètre ou input X1, tandis que la partie Y agit sur X en réalisant le paramètre ou input Y2. Par conséquent, si l'on se fie aux tableaux qui définissent le comportement de ces parties, la partie X et la partie Y changeront en B et en I respectivement. Le système Z aura donc comme première transition d'état (A, J) → (B, I). Ensuite, l'état (B, I) va déterminer de nouveaux inputs, c'est-à-dire les inputs X2 et Y1. En vertu de ces inputs et du fait que le système est dans l'état (B, I), on peut ensuite inférer, par la même démarche qui nous a permis d'inférer la première transition d'état, que le système Z va ensuite passer de l'état (B, I) à l'état (D, K). Nous pourrions continuer de la sorte et, en principe, déterminer l'ensemble des transitions d'état du système Z ou, autrement dit, son comportement.

Le comportement (l'ensemble des transitions d'état) qui découle de ce mécanisme est de type déterministe. On s'en convainc quand on constate que, pour chacun des états du système Z, ce mécanisme ne produit qu'une seule transition d'état. Par exemple, à partir de l'état (A, J), ce mécanisme permet seulement la transition (A, J) → (B, I). De plus, pour connaître le comportement futur du système, il suffit de connaître l'état actuel, disons (A, J); la connaissance des états antérieurs à (A, J) n'est pas nécessaire pour savoir que le système va ensuite passer à l'état (B, I).

Maintenant que nous en connaissons davantage sur le type de mécanisme qui est à la base du comportement des systèmes dynamiques, il est temps de passer à la prochaine étape, soit d'établir le mécanisme d'apprentissage qui permet au système moteur de passer d'une exécution inadéquate du mouvement à une exécution adéquate du mouvement.

7.2. Le mécanisme d'apprentissage inspiré de Ashby

Comme nous l'avons dit plus haut, c'est en changeant le poids de ses connexions que le réseau de neurones ou contrôleur parvient à émettre des commandes motrices appropriées, c'est-à-dire des commandes motrices qui causent un mouvement qui redonne à la variable de contrôle la valeur souhaitée (la valeur qui spécifie le succès). Le premier élément du mécanisme d'apprentissage est composé de ce réseau de neurones ainsi que de la synergie que le réseau de neurones contrôle. Je nomme ce système « système R », la lettre « R » étant mise pour le terme réactif[64]. Afin de définir complètement notre mécanisme d'apprentissage, cinq composantes doivent être ajoutées. J'introduis au cours des quatre sous-sections suivantes ces cinq composantes ainsi que les relations causales qui existent entre elles.

7.2.1. Le système R interagit avec l'environnement (les variables optiques)

Selon notre modèle de l'action, la perception d'une variable optique agit causalement sur le système R. Et, si l'on recule un peu plus tôt dans la chaîne causale, c'est la variable optique qui agit causalement sur le système perceptuel. Donc, par transitivité, l'environnement (la variable optique) agit causalement sur le système R. Mais l'environnement (la variable optique) n'est pas le seul à avoir un impact causal; le système R a aussi un impact causal sur l'environnement. En effet, à travers nos mouvements, nous modifions la valeur de la variable optique. Nous pouvons donc poser l'existence d'une interaction entre le système R et l'environnement (la variable optique). L'autre élément du mécanisme d'apprentissage de Ashby est précisément l'environnement (la variable optique) et, dans ce mécanisme d'apprentissage, ces deux parties (la variable optique et le système R)[65] interagissent causalement (voir figure 7.1).

---

[64] Le réseau de neurones *réagit* aux variables optiques.

[65] À ma connaissance, Ashby ne prend pas la peine de définir précisément le système R. C'est donc moi qui le définis comme un contrôleur par feedback. Et quant à l'environnement, Ashby le définit comme un environnement égocentrique; il ne fait pas référence à une variable optique de contrôle, au sens où cette dernière est comprise par la tradition écologiste.

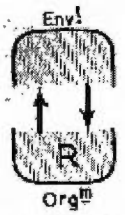


Figure 7.1 L'interaction entre le système R et l'environnement. Tirée de Ashby (1960).

7.2.2. La paramétrisation du système R

Pour développer son mécanisme d'apprentissage, Ashby (1960) met en scène un chaton qui veut se réchauffer auprès d'un feu. Dans cet exemple, on remarque que s'il est confronté pour la première fois au feu, le chaton se déplace de manière inappropriée avant de se déplacer de manière appropriée. Par exemple, avant de se placer à une distance adéquate du feu, le chaton pourrait s'avancer exagérément vers lui (un premier mouvement), s'en éloigner exagérément (un deuxième mouvement), voire tenter d'attraper des tisons enflammés (un troisième mouvement). Nous pourrions également prendre comme exemple un joueur de tennis qui pratique son coup droit à l'aide d'une distributrice de balles qui lui projette des balles. À chaque balle envoyée, le joueur modifie sensiblement le coup droit qu'il exécute, cela jusqu'au moment où il trouve une exécution appropriée. Maintenant, le problème est de comprendre ces changements moteurs (ces changements de mouvement).

Tout changement de mouvement provient du fait que les paramètres du système R changent de valeur. Les paramètres du système R sont de deux sortes. On compte d'abord des paramètres de type externe. Ces paramètres correspondent à l'environnement. Or, les exemples que nous avons donnés montrent que l'exécution d'un nouveau mouvement ne provient pas d'un changement au sein de l'environnement. En effet, dans nos deux exemples, celui du chaton et celui du joueur de tennis qui pratique son coup droit, l'environnement reste identique. Donc, les paramètres qui expliquent le changement moteur doivent être internes au système R. Or, les paramètres internes du système R (un réseau de neurones) sont le poids des connexions du réseau de neurones (le système R). Donc, le changement moteur est le résultat d'une pondération différente des connexions du réseau de neurones. On peut donc ajouter les paramètres internes (le poids des connexions) du système R à notre mécanisme

d'apprentissage. Ces paramètres internes agiraient causalement sur le système R. Dans la figure ci-dessous, les paramètres internes du système R sont représentés par le symbole S.

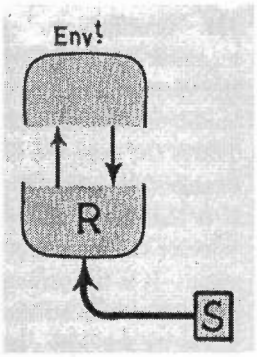


Figure 7.2 L'action des paramètres S sur le système R. Tirée de Ashby (1960).

C'est donc à la suite d'un changement d'état du système S (à la suite d'un changement du poids des connexions) que le système R donne une autre valeur aux paramètres de contrôle de la structure coordinative, cela lorsque le système R est affecté par la même variable optique. Maintenant, tout le problème est de comprendre comment le poids des connexions (le système S) peut être changé de manière à ce que le contrôleur moteur émette des commandes motrices qui vont produire un mouvement ayant la particularité d'éliminer l'erreur spécifiée par la variable optique de contrôle, c'est-à-dire redonner à la variable de contrôle la valeur souhaitée. Dorénavant, lorsque je parlerai d'une pondération appropriée du réseau de neurones ou contrôleur, j'aurai en tête une pondération qui permet de réaliser ce qui précède, c'est-à-dire redonner à la variable de contrôle la valeur souhaitée. Le mécanisme d'apprentissage que je vais maintenant décrire permet d'expliquer comment le réseau de neurones ou contrôle acquiert cette pondération appropriée.

7.2.3. La variable d'appréciation

Le troisième élément du mécanisme d'apprentissage est la variable d'appréciation[66]. La variable d'appréciation correspond à la différence qui existe entre, d'une part, l'état de

[66] Ashby parle non pas d'une variable d'appréciation, mais de variables essentielles. Les variables essentielles chez Ashby désignent les variables qui sont responsables de la survie biologique de

l'environnement qui accompagne le mouvement et, d'autre part, l'état environnemental auquel réfère l'affordance. Plus précisément, les choses se passent à peu près ainsi. Lorsque le mouvement est exécuté, l'environnement change d'état. L'environnement, ici, ne désigne pas seulement les objets qui sont extracorporels; il comprend aussi le corps de l'agent. Dans ces circonstances, on comprend que l'exécution d'un mouvement est nécessairement accompagnée par un changement environnemental. L'agent prend connaissance de ce changement environnemental grâce à ses sensations visuelles ou proprioceptives. Donc, via ses sensations visuelles ou proprioceptives, l'agent possède de l'information sur ce nouvel état environnemental. Cette information est alors comparée à l'état de l'environnement auquel fait référence l'affordance. Maintenant, il peut exister une différence entre l'état de l'environnement qui fait suite au mouvement et l'état de l'environnement auquel fait référence l'affordance. Cette différence, qui peut être nulle, correspond à la variable d'appréciation. Si cette différence est nulle, le joueur sera satisfait de son mouvement; si elle est plutôt grande, insatisfait. On peut supposer que le joueur possède des limites à l'intérieur desquelles la variable d'appréciation sera « bonne » et au-delà desquelles le coup sera jugé ou apprécié comme mauvais. C'est cette variable d'appréciation ainsi que ces limites d'appréciation qui font partie du mécanisme d'apprentissage. Comme on vient de le voir, c'est l'état de l'environnement qui agit causalement sur cette variable. Nous pouvons donc ajouter à notre mécanisme d'apprentissage une relation causale allant de l'environnement vers la variable d'appréciation (voir figure 7.3).

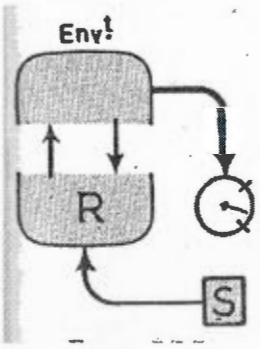


Figure 7.3 L'ajout de la variable d'appréciation. Tirée de Ashby (1960).

---

l'organisme. Comme on va le voir, la variable d'appréciation est très différente de ce type de variable. Je me démarque donc d'Ashby sur ce point.

La figure 7.3 ne reflète pas fidèlement les propos qui viennent d'être tenus. Selon ce qui a été dit, la valeur que prend la valeur d'appréciation n'est pas seulement fonction de l'environnement, mais est aussi fonction de l'affordance. Nous devrions donc ajouter un autre élément à la figure 7.3, soit l'affordance. Et il faudrait poser une relation causale allant de cette affordance vers la variable d'appréciation. Je ne représenterai pas ce nouvel élément dans une nouvelle figure. Il suffit de se l'imaginer.

7.2.4. L'action de la variable d'appréciation sur les paramètres S

Le chaton n'arrive pas du premier coup à se déplacer de manière adéquate, c'est-à-dire à se placer à une distance adéquate du feu. Et, lorsqu'il est confronté aux balles de la distributrice de balles, le joueur de tennis ne parvient pas immédiatement à exécuter son coup droit de manière adéquate. Avant d'exécuter un mouvement réussi, on observe que le système R exécute plusieurs mouvements plus ou moins réussis. Bref, on observe qu'il procède par essais et erreurs. Il importe ici de préciser ce que l'on veut dire par « procéder par essais et erreurs ». On veut dire par là que le système R exécute des mouvements et que ceux-ci sont successivement rejetés, cela jusqu'au moment où l'un d'eux ne met plus en « danger » la variable d'appréciation ou, plus précisément, jusqu'au moment où le mouvement n'a plus pour effet de « déplacer » cette variable en dehors de ses limites. Ce dernier mouvement est alors conservé. Ainsi, dire que le système R procède par essais et erreurs, c'est dire que : 1) si la variable d'appréciation est en dehors de ses limites, le système R exécute un autre mouvement; et 2) si la variable d'appréciation est à l'intérieur de ses limites, le système R conserve son mouvement. Les mouvements du système R dépendent donc de la valeur de la variable d'appréciation. L'effet de la variable d'appréciation sur le mouvement du système R n'est cependant pas direct. Ce sont les paramètres « S » qui influent directement sur le mouvement initié par le système R. Par conséquent, pour déterminer le mouvement du système R, la variable d'appréciation doit agir sur les paramètres « S ». Nous pouvons donc rajouter l'élément suivant au mécanisme d'apprentissage (voir figure 7.4).

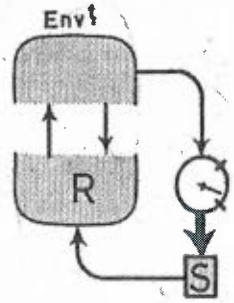


Figure 7.4 L'action de la variable d'appréciation sur les paramètres S. Tirée de Ashby (1960).

Avant de continuer, il faudrait être plus explicite à propos de l'action de la variable d'appréciation sur les paramètres « S ». Nous pouvons préciser l'action de la variable d'appréciation sur les paramètres « S » en tenant compte de la manière dont les mouvements du système R dépendent de la variable d'appréciation. Ci-dessus, en nous penchant sur le processus d'essais et erreurs, nous avons remarqué deux choses à propos de la dépendance que les mouvements entretiennent à l'endroit de la variable d'appréciation : 1) si la variable d'appréciation n'est pas à l'intérieur de ses limites, un nouveau mouvement est entrepris; 2) si elle est à l'intérieur de ses limites, le mouvement est maintenu. Afin de préciser l'action de la variable d'appréciation sur les paramètres « S », nous allons montrer comment la variable d'appréciation doit agir sur les paramètres « S » pour que 1) et 2) soient vrais.

Pour que 1) soit vrai, l'action de la variable d'appréciation sur les paramètres « S » doit prendre la forme suivante : si la variable d'appréciation n'est pas à l'intérieur de ses limites, l'état de « S » n'est pas un état d'équilibre. En définissant de la sorte l'action de la variable d'appréciation « hors limites » sur le sous-système « S », il est clair que l'observation 1) est respectée. En effet, si les paramètres S ne sont pas en équilibre, ils vont changer de valeur et, s'ils changent de valeur, le système R va adopter un autre mouvement.

Pour que 2) soit vrai, l'action de la variable d'appréciation sur les paramètres « S » doit prendre la forme suivante : si la variable d'appréciation est à l'intérieur de ses limites, l'état de « S » correspond à un état d'équilibre. De la sorte, le mouvement qui ne pousse pas la variable d'appréciation en dehors de ses limites est maintenu. En effet, si S est en équilibre, il maintient sa valeur et, s'il maintient sa valeur, le système R maintiendra son mouvement.

Jusqu'à présent, nous avons présenté les cinq composantes du mécanisme d'apprentissage ainsi que les relations causales entre ces composantes[67]. Or, ces composantes et leurs relations causales ont été exposées sans qu'on tienne compte de la dynamique qui existe entre elles. À ce titre, notre compréhension du mécanisme d'apprentissage est plutôt morcelée et, à vrai dire, très peu intuitive. Remédions à cette situation et montrons comment les différents éléments de ce mécanisme s'articulent entre eux.

Tout d'abord, un mouvement est exécuté par le système R. Ce mouvement change l'état de l'environnement, et ce nouvel état de l'environnement affecte la variable d'appréciation. Si l'environnement déplace la variable d'appréciation à l'extérieur de ses limites, cette variable va agir sur les paramètres « S » de manière à ce que ces derniers changent de valeur. Cela découle du fait que, lorsque la variable d'appréciation est en dehors de ses limites, le système « S » est en déséquilibre. Ce changement de valeur des paramètres « S » va instaurer une nouvelle loi de contrôle et, du fait même, un nouveau mouvement sera produit par le système R. Si ce nouveau mouvement confronte de nouveau l'organisme à un environnement qui « déplace » la variable d'appréciation en dehors de ses limites, le cycle qui vient d'être décrit va recommencer. Ce cycle ne peut se terminer qu'à une seule condition : si le mouvement produit par le système R cause un état de l'environnement qui permet à la variable d'appréciation de rester dans ses limites. C'est à cette condition que ce cycle se termine, car lorsque la variable d'appréciation est à l'intérieur de ses limites, les paramètres « S » conservent leur valeur et, du fait même, le mouvement (la loi de contrôle) est conservé. À ce moment, le système R a exécuté un mouvement qui satisfait l'affordance, puisque, comme on l'a dit, la variable d'appréciation est à l'intérieur de ses limites si, et seulement si, l'état environnemental (visuel ou proprioceptif) accompagnant le mouvement correspond à l'état environnemental auquel réfère l'intention.

Mais la pondération du contrôleur moteur est-elle appropriée? Autrement dit, permet-elle de produire un mouvement qui ramène la variable de contrôle à la valeur souhaitée? Nous venons de dire que, à la suite du processus d'apprentissage, le mouvement exécuté par le système R satisfaisait la perception de l'affordance. Or, si la perception de l'affordance est satisfaite, alors la variable de contrôle a la valeur souhaitée. Nous pouvons donc conclure que

[67] La cinquième composante, qui n'est pas représentée dans la figure 7.4, est l'affordance, qui agit causalement sur la variable d'appréciation.

le mécanisme d'apprentissage décrit ci-dessus permet de donner au réseau de neurones (le contrôleur) une pondération appropriée.

Avant de passer au prochain chapitre, je vais montrer que le mécanisme d'apprentissage qui vient d'être décrit est en mesure de répondre à une objection importante.

Le mécanisme d'apprentissage qui a été décrit comprend l'apprentissage comme une suite d'essais et d'erreurs. Or, apprendre par essais et erreurs peut être problématique. Cela peut être problématique, car, lorsqu'on procède par essais et erreurs, les essais peuvent ne pas s'approcher ou tendre vers l'essai réussi. Autrement dit, lorsque l'on procède par essais et erreurs, on procède de manière aléatoire; la succession de nos essais ne se caractérise pas nécessairement par un progrès. Cela n'est pas grave lorsque le champ des essais possibles est limité. Mais il en va autrement lorsque les différents essais sont en nombre quasi infini et que, au sein de ce champ des possibles, les essais réussis sont en très petit nombre. Dans ces circonstances, l'apprentissage peut ne jamais prendre fin. De plus, des considérations empiriques semblent réfuter cette progression aléatoire des essais : un examen empirique révèle que les essais suivent une certaine direction — qu'ils ne sont pas effectués de manière purement aléatoire (Merleau-Ponty, 1942).

Notre mécanisme d'apprentissage est-il en mesure de répondre à cette objection? Peut-il faire en sorte que la succession des essais n'ait pas un caractère purement aléatoire? Plus précisément, ce mécanisme d'apprentissage peut-il nous assurer que le système S (le poids du réseau de neurones) va recevoir une valeur appropriée (une pondération appropriée) à la suite d'un processus qui n'est pas purement aléatoire? Absolument.

Le système R, l'environnement, le système composé de la variable d'appréciation et le système S forment les parties d'un système englobant. Plus précisément, l'environnement et le système R forment un sous-système, la variable d'appréciation et le système S forment un autre sous-système, et ces deux sous-systèmes interagissent (autrement dit, il y a du feedback entre eux), ce qui fait en sorte qu'ils forment un système englobant. Nommons ce système englobant le système T. Notre question est donc celle-ci : sommes-nous assurés que le système T évolue de manière à donner au système S une valeur appropriée (c.-à-d. une pondération appropriée, qui ramène la variable de contrôle à la valeur souhaitée)? *Si le*

système T est stable, nous sommes effectivement assurés que les paramètres S recevront une valeur appropriée. C'est ce que je vais tenter de montrer dans ce qui suit.

Pour montrer que l'évolution du système T est telle que les paramètres S vont recevoir une valeur appropriée, je dois d'abord introduire la notion de « step-function ».

Une « step-function » correspond au comportement typique d'une variable ou d'un système. Lorsqu'une variable peut être représentée par une « step-function », cette variable se caractérise par des intervalles finis de constance séparés par des sauts instantanés. Par exemple, si un morceau de caoutchouc est étiré, la force que le morceau de caoutchouc exerce est approximativement proportionnelle à la longueur de l'étirement. La constante de proportionnalité conserve une valeur constante, cela jusqu'au moment où le caoutchouc cède. À ce moment, la constante de proportionnalité change spontanément de valeur; elle acquiert spontanément une valeur de zéro. Cette variable (la constante de proportionnalité) se comporte donc comme une « step-function ». De même, si un acide puissant est ajouté de manière continue à une solution alcaline, le pH garde la même valeur et change spontanément de valeur. Le pH est donc une variable qui peut être représentée par une « step-function ». Maintenant, qu'est-ce qui fait qu'une variable pouvant être représentée par une « step-function » change spontanément de valeur? Ce changement est fonction de la valeur que prennent les variables dont dépend la variable « step-function ». Si ces autres variables — appelons-les « variables principales » — atteignent une *valeur critique*, la variable « step-function » changera de valeur instantanément. La figure suivante représente un système ayant une variable « step-function » (S) et deux variables principales (A et B). On remarque que si les variables principales atteignent une valeur critique (la ligne C), S prend une autre valeur et le portrait de phase correspondant aux variables principales A et B change de forme.

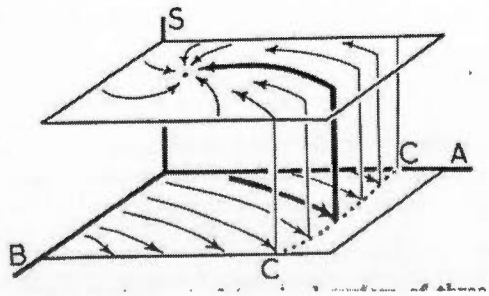


Figure 7.5. Portrait de phase d'un système ayant une variable « step-function ». Tiré de Ashby (1960).

Maintenant, suivant Ashby (1960), mettons ces considérations sur les « step-functions » en relation avec le système T. Nous allons poser que le système S (le poids des connexions du réseau de neurones) est une variable qui peut être représentée par une « step-function » et nous allons poser que le système R, l'environnement et la variable d'appréciation sont les variables principales. De plus, nous allons poser que la valeur critique des variables principales correspond aux valeurs où la variable d'appréciation est à l'extérieur de ses limites. À l'aide de la figure 7.5, on peut maintenant montrer que si le système T est stable, le système S prendra une valeur appropriée.

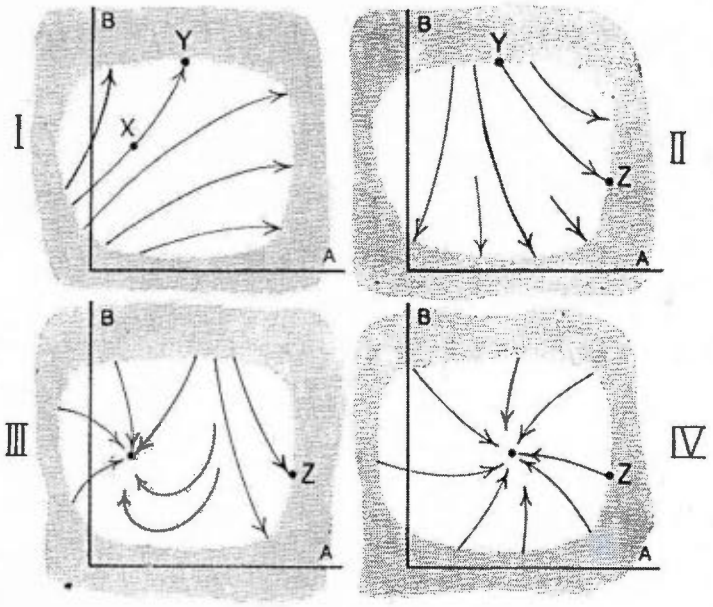


Figure 7.6. Les portraits de phase correspondant aux différentes valeurs du système S. Tiré de Ashby (1960).

Pour simplifier la description, les portraits de phases de la figure 7.6 représentent le comportement, non pas de trois variables (système R, environnement et variable d'appréciation), mais de deux variables. Et nous posons que la région ombragée des portraits de phase correspond à des états où la variable d'appréciation est à l'extérieur de ses limites. Autrement dit, la région ombragée correspond aux valeurs critiques. Dans le premier portrait de phase, le système S (la variable « step-function ») prend la valeur I. Le premier portrait de phase découle de l'adoption de cette pondération. Dans le premier portrait de phase, les

variables principales se trouvent d'abord au point X. Le point X représente l'état du système R, l'état de l'environnement et l'état de la variable d'appréciation. Ces trois variables évoluent conformément au portrait de phase I. La trajectoire sur laquelle se trouve le point X mène vers le point Y. Or, le point Y est un point où la variable d'appréciation est en dehors de ses limites. Plus précisément, cette trajectoire signifie que, conséquemment à cette pondération (correspondant à la valeur I), le système R et l'environnement atteignent un point où la variable d'appréciation est en dehors de ses limites. Les variables principales atteignent donc un point critique, le système S va changer de valeur. Autrement dit, le poids des connexions du réseau de neurones va spontanément changer de valeur, il va prendre la valeur II. Un nouveau portrait de phase prendra donc place. Comme ce nouveau portrait de phase n'est pas plus stable que le précédent, les variables principales vont changer de manière à adopter la valeur Z. Comme le point Z correspond à un point critique, le système S va de nouveau changer de valeur (la valeur III) et un nouveau portrait de phase prendra place. Compte tenu de ce troisième portrait de phase, le point Z ne mène pas vers un point où la variable d'appréciation est à l'intérieur de ses limites; la variable S changera de nouveau de valeur et prendra la valeur IV. Comme ce portrait de phase est stable, les variables principales vont atteindre un point d'équilibre. Avec cette nouvelle valeur du système S, peu importe la valeur initiale des variables principales, les variables principales vont toujours mener vers un point où la variable d'appréciation est à l'intérieur de ses limites. À ce titre, la pondération du système S satisfait la perception de l'affordance. Cette pondération ramène donc la variable de contrôle à la valeur souhaitée et est donc une pondération appropriée.

Il faut voir ici que nous supposons que le système T est stable. Cela signifie que, peu importe l'écart qui sépare l'état initial de ce système et l'état qui correspond à un état d'équilibre, ce système se comportera toujours de manière à trouver cet état d'équilibre. Plus précisément, le système T se comporte toujours de manière à trouver un état du système S qui conditionne un portrait de phase stable des variables principales. Donc, les différentes valeurs que prend le système S sont orientées vers cet état d'équilibre; ces valeurs ne sont pas déterminées aléatoirement.

Pour finir, il faut se demander pourquoi un point d'équilibre du système T correspond toujours à un point où la variable d'appréciation est à l'intérieur de ses limites. L'argument

qui est à la base de cette correspondance est simple. Si le système T est à l'équilibre, le système S est à l'équilibre. Or, par définition, si le système S est à l'équilibre, la variable d'appréciation est à l'intérieur de ses limites.

Dans la présentation qui a été donnée de notre mécanisme d'apprentissage, la variable optique était une variable optique appropriée, c'est-à-dire une variable optique qui permet de satisfaire la perception de l'affordance. Ainsi, si le joueur de baseball perçoit l'affordance « attrape-moi » présentée par la balle de baseball frappée au champ centre, l'agent perçoit ensuite la variable optique qui permet de satisfaire cette perception d'affordance, c'est-à-dire la variable optique « accélération de l'angle d'élévation ». Mais que se passe-t-il avec notre mécanisme d'apprentissage lorsque l'agent ne perçoit pas la variable optique appropriée? Il arrive alors que le système T ne se stabilise pas et donc, que les paramètres S changent sans cesse de valeur. On comprend pourquoi le système T ne se stabilise pas. Si la variable optique est inappropriée, elle ne spécifie pas le succès ou l'échec du mouvement auquel réfère l'affordance. Or, si la variable optique ne spécifie pas le succès ou l'échec du mouvement auquel réfère l'affordance, l'agent ne peut pas se servir de cette variable optique pour causer le mouvement auquel réfère l'affordance. Donc, si la variable optique est inappropriée, cette variable optique ne peut pas causer le mouvement auquel réfère l'affordance. Or, si cette variable optique ne cause pas le mouvement auquel réfère l'affordance, le mouvement qu'elle cause fait en sorte que la variable d'appréciation reste en dehors de ses limites. Donc, si la variable optique est inappropriée, la variable d'appréciation restera en dehors de ses limites. Or, si la variable d'appréciation est en dehors de ses limites, les paramètres S vont changer. Donc, si la variable optique est inappropriée, les paramètres S seront toujours en changement. Afin de remédier à cette impasse, l'agent va devoir percevoir une autre variable optique. Comment l'agent apprend-t-il à percevoir la variable optique appropriée? C'est ce que nous allons tenter de voir dans le prochain chapitre. Plus précisément, dans le prochain chapitre, nous allons voir comment l'agent apprend à percevoir l'affordance appropriée et la variable optique appropriée.

7.3 Résumé

La variable optique peut être appropriée ou non. Elle est appropriée lorsqu'elle peut causer et contrôler le mouvement auquel réfère l'affordance. Dans ce chapitre, nous avons supposé que la variable optique était appropriée.

Le système R (le réseau de neurones) interagit avec la variable optique. La façon dont le système R interagit avec l'environnement dépend de la pondération du système R. Si le système R interagit mal avec l'environnement, la pondération est inappropriée; sinon, elle est appropriée. Donc, l'apprentissage moteur consiste à changer le poids des connexions (le système S) du système R, cela de façon à ce que le système R donne aux paramètres de contrôle de la structure coordinative une valeur qui permette de réduire à zéro l'erreur spécifiée par la variable optique. Notre problème, dans ce chapitre, a précisément été de trouver un mécanisme qui permette de donner un poids approprié aux connexions du réseau de neurones, c'est-à-dire un poids qui fera en sorte que les paramètres de contrôle de la structure coordinative vont réduire à zéro l'erreur spécifiée par la variable optique.

Nous avons résolu notre problème en connectant le système S (le poids des connexions du réseau de neurones) à une variable d'appréciation, qui est elle-même connectée à l'environnement et à l'affordance. L'idée de base est la suivante. Lorsque l'état de l'environnement (visuel ou proprioceptif) est suffisamment différent de l'état auquel réfère l'affordance, la variable d'appréciation est en dehors de ses limites, et cela fait en sorte que le système S n'est pas dans un état d'équilibre : il change alors d'état. Le système R interagit donc différemment avec l'environnement. Un autre mouvement est exécuté, ce qui se traduit par un nouvel environnement (visuel ou proprioceptif). Si ce nouvel environnement est de nouveau différent de l'état auquel réfère l'affordance, la variable d'appréciation est de nouveau en dehors de ses limites, ce qui cause de nouveau le même cycle. Ce cycle prend fin lorsque le nouvel état de l'environnement, causé par le nouveau comportement moteur, est proche de l'état auquel réfère l'affordance. En effet, lorsque c'est le cas, la variable d'appréciation est à l'intérieur de ses limites, ce qui fait en sorte que les paramètres S restent dans le même état; le système S est à l'équilibre. La pondération appropriée du réseau de neurones a été trouvée.

CHAPITRE VIII

L'APPRENTISSAGE D'UNE BONNE POLITIQUE PERCEPTUELLE

Dans ce chapitre, mon objectif est de décrire le mécanisme d'apprentissage qui permet de passer d'un agent qui ne fait pas attention aux bonnes affordances (ni aux bonnes variables optiques) à un agent qui fait attention aux bonnes affordances (et aux bonnes variables optiques)[68]. Cet apprentissage perceptuel n'est pas supervisé. En effet, tout au plus, le feedback de l'instructeur porte sur les plans (les intentions) de l'agent; pas sur ses perceptions. L'agent doit donc apprendre par lui-même à faire attention aux affordances et aux variables optiques qui peuvent lui servir. J'ai trouvé chez Sutton et Barto (1998) un mécanisme d'apprentissage non supervisé. Le mécanisme d'apprentissage présenté par Sutton et Barto est qualifié d'apprentissage par renforcement. Dans ce chapitre, je veux montrer comment l'apprentissage par renforcement permet à l'agent de faire attention aux bonnes affordances et aux bonnes variables optiques. Il existe au moins deux méthodes d'apprentissage par renforcement. Je vais présenter ces deux méthodes, ainsi qu'une troisième méthode, qui n'est pas à proprement parler une méthode d'apprentissage par renforcement, mais dont l'exposition permet de bien comprendre les deux premières méthodes. Ces trois méthodes se rejoignent sur ce point : l'agent apprend à percevoir les bonnes affordances et les bonnes variables optiques parce qu'il cherche à maximiser ses récompenses.

Avant de commencer, j'aimerais faire la précision suivante. L'apprentissage par renforcement qui sera présenté dans ce chapitre concerne autant l'apprentissage qui est à la base de la perception de la bonne affordance que l'apprentissage qui est à la base de la

---

[68] Plus tard, le concept d'attention sera défini. Pour l'instant, on peut se contenter du concept ordinaire d'attention, selon lequel l'attention est un acte de concentration consciente sur un objet.

perception de la bonne variable optique. Cependant, afin de simplifier mon exposé, le développement qui suit concerne plus particulièrement l'apprentissage de la bonne perception d'affordance. À la fin du chapitre, je prendrai la peine de montrer que les propos tenus au cours du développement du présent chapitre s'appliquent également à l'apprentissage qui est à la base de la perception de la bonne variable optique.

8.1 Le problème de l'apprentissage par renforcement

Au cours d'un épisode d'actions, c'est-à-dire d'une séquence d'actions élémentaires qui possède un début et une fin, l'agent est en constante interaction avec son environnement. L'agent exécute une action et cette action cause un changement d'état de l'environnement. Ce nouvel état de l'environnement cause ensuite une nouvelle action chez l'agent, qui cause un nouvel état de l'environnement, etc. C'est ainsi que l'agent et l'environnement interagissent. Lorsque l'agent exécute une action et que cette action cause un nouvel état de l'environnement, ce nouvel état de l'environnement peut être accompagné d'une récompense (positive ou négative). La récompense qui accompagne le nouvel état de l'environnement est nommée « récompense immédiate ». Mais le nouvel état de l'environnement n'apporte pas nécessairement avec lui une récompense immédiate. Il est possible que, au cours d'un épisode, seuls des états situés plus loin dans le temps, c'est-à-dire des états causés par des actions subséquentes, apportent des récompenses à l'agent. Le but de l'agent est de maximiser les récompenses qu'il recevra tout au long de l'épisode; donc, pas seulement les récompenses immédiates.

C'est en passant par un processus d'apprentissage que l'agent parvient à atteindre son but (maximiser ses récompenses). En effet, le processus d'apprentissage modifie les actions de l'agent de manière à ce que celles-ci rapportent davantage de récompenses. Les actions de l'agent sont fonction de sa *politique*. La politique de l'agent est un « mapping » des états de l'environnement sur les actions de l'agent. Et lorsque je dis que les actions de l'agent sont fonction de sa politique, je veux simplement dire ceci : si l'agent possède une politique où l'état X de l'environnement est associé à (ou « mappé » sur) l'action Y et si l'état X de l'environnement est présent, alors l'agent exécutera l'action Y. Donc, comme les actions de l'agent sont fonction de sa politique et comme l'apprentissage a pour but d'amener l'agent à

produire des actions qui rapporte un maximum de récompenses, le processus d'apprentissage doit modifier la politique de l'agent de manière à constituer une politique optimale, c'est-à-dire une politique dont les actions vont rapporter un maximum de récompenses. Dans l'apprentissage par renforcement, la modification de la politique passe non pas par le feedback de l'instructeur, mais par des interactions avec l'environnement. Le problème de l'apprentissage par renforcement est donc celui-ci : Comment, à partir d'interactions avec son environnement, l'agent peut-il changer sa politique de manière à la rendre optimale?

Avant de présenter les méthodes d'apprentissage par renforcement qui résolvent ce problème, il faudrait apporter des précisions sur certaines des notions qui viennent d'être introduites. En particulier, les notions d'agent, d'état environnemental et d'action méritent d'être examinées.

Un agent est un système qui prend des décisions ayant un effet sur son environnement. Et l'environnement est compris ici comme la source d'inputs du système ou agent. L'ensemble des agents est plus vaste qu'il n'y paraît. Par exemple, Sutton et Barto (1998, p. 54) considèrent l'instance de contrôle d'un bioréacteur comme un agent. Cette instance de contrôle prend des décisions sur la température et le brassage à imposer à la cuve remplie d'éléments biochimiques, et cette cuve, via certains senseurs, est une source d'inputs pour cette instance de contrôle. Les organismes sont aussi des agents. En effet, les organismes prennent des décisions motrices qui agissent sur l'environnement physique, qui est une source d'inputs pour l'organisme. Mais il y a plus. Si l'on se fie à notre définition de l'agent (un système qui prend des décisions ayant un effet sur son environnement) et de l'environnement (une source d'input pour l'agent), l'organisme peut être lui-même composé d'agents. Plus précisément, il n'est pas impossible que les différentes fonctions cognitives de l'organisme doivent elles-mêmes être considérées comme des agents. En effet, une fonction cognitive émet des outputs et ces outputs peuvent être considérés comme des décisions ayant un effet sur un environnement, c'est-à-dire sur une source d'inputs pour la fonction cognitive en question. Et cet environnement n'est pas nécessairement l'environnement physique. Par exemple, la fonction cognitive peut prendre une décision qui a un effet sur une autre fonction cognitive et c'est cette deuxième fonction cognitive qui doit être considérée comme

l'environnement de la première fonction cognitive; les outputs de cette deuxième fonction cognitive seraient alors les inputs de la première fonction cognitive[69].

Ces considérations nous amènent évidemment à élargir les notions d'état environnemental et d'action. Selon ce qui vient d'être dit, les états environnementaux ne sont pas seulement les états de l'environnement physique. Comme on l'a dit, un état de l'environnement est tout ce qui peut servir d'input à l'agent. Donc, en plus de pouvoir correspondre à des états de l'environnement physique, les états de l'environnement pourraient correspondre, par exemple, à des états purement psychologiques (comme un état de surprise) ou à un mixte des deux (la mémoire d'une expérience sensible, par exemple). Il suffit qu'un système ait comme input ces états psychologiques pour que ces derniers soient considérés comme des états environnementaux.

De même, les actions peuvent référer à plusieurs choses différentes. Sutton et Barto (1998) définissent l'action comme une *décision* pouvant affecter l'environnement. Notons au passage que cette définition de l'action ne correspond pas à la définition de l'action qui a été donnée au chapitre 1, où l'action était comprise comme la production d'un mouvement par une cause interne. Pour bien comprendre le présent chapitre, il faut faire abstraction de la définition de l'action qui a été donnée au chapitre 1. Afin de bien distinguer notre concept original d'action (le concept dretskien d'action) de ce dernier concept d'action, je vais placer un astérisque à la suite du mot « action » à chaque fois que ce mot réfère au dernier concept d'action, c'est-à-dire à chaque fois que ce mot réfère à une décision pouvant avoir un effet sur l'environnement de l'agent. Ceci étant dit, revenons maintenant à notre propos initial et montrons que les actions* peuvent référer à des choses bien différentes. Les actions* peuvent bien sûr correspondre aux commandes (ou décisions) motrices qui sont envoyées aux membres du corps. Mais elles peuvent également correspondre à des décisions purement mentales. Par exemple, une action* pourrait être une décision qui détermine ce que l'agent doit croire ou une décision qui détermine la direction que l'attention de l'agent doit prendre.

---

[69] Et il faut voir que, selon notre description, la deuxième fonction cognitive est aussi un agent puisque ses outputs agissent sur la première fonction et que les outputs de la première fonction sont des inputs pour la deuxième fonction. Nous avons donc ici deux agents qui interagissent entre eux et dont chacun est l'environnement de l'autre.

Cinq notions ont été dégagées de notre analyse du problème de l'apprentissage par renforcement. Plus précisément, nous avons dégagé les notions d'agent, d'action*, d'état environnemental, de récompense et de politique. Essayons maintenant de déterminer à quoi correspondent ces notions lorsqu'elles sont appliquées à notre propre problème.

Pour nous, l'agent est le système perceptuel. C'est ce système qui pose des actions* sur la base d'une politique, et c'est lui qui, à travers un apprentissage par renforcement, doit trouver une politique optimale. Définissons maintenant l'environnement de cet agent, ses actions*, ses récompenses et sa politique.

L'état de l'environnement à un moment donné correspond pour nous à l'ensemble des affordances comprises dans l'environnement à un moment donné. Donc, pour nous, les états de l'environnement ne sont pas des états purement psychologiques, mais, au contraire, des états physiques.

L'action* de l'agent (ou système perceptuel) correspond à la décision de l'agent (le système perceptuel) de diriger son attention vers l'une des affordances présentées par l'état de l'environnement. Ainsi, si l'état de l'environnement me présente les affordances x, y et z, l'action* du système perceptuel (l'agent) pourrait correspondre à la décision de faire attention à l'affordance x. Maintenant, il faut prendre garde au terme « attention ». D'habitude, l'attention est comprise comme un phénomène conscient, d'ordre personnel (par opposition à sous-personnel). Ce n'est pas nécessairement le cas pour nous. Il arrive que l'agent « fasse attention » à des affordances sans en avoir conscience. C'est ce qui se passe avec la plupart de nos comportements automatiques. Afin de rendre compte des perceptions d'affordances qui sont impliquées dans ces comportements automatiques, je vais poser que l'expression « faire attention à » signifie simplement « tenir compte de » ou « est sensible à ». Ainsi, lorsque je dis que le système perceptuel fait attention à une affordance, je me contente de dire qu'il est sensible ou qu'il tient compte de cette affordance à l'exclusion des autres affordances de son environnement.

Maintenant que les états environnementaux et les actions* ont été définis, il est facile de définir la politique du système perceptuel (l'agent). La politique, comme on l'a dit, correspond à un « mapping » des états environnementaux sur les actions*. Pour nous, un état environnemental est un ensemble de propriétés perceptibles (comprenant des affordances), et

une action* est une décision de faire attention à une affordance plutôt qu'aux autres. Donc, pour nous, la politique est un « mapping » d'ensembles comprenant des propriétés perceptibles (y compris des affordances) sur des directions qui sont données à l'attention. Ce « mapping », comme on l'a dit plus haut, implique ceci : lorsqu'un état particulier de l'environnement se produit, c'est-à-dire lorsqu'un ensemble d'affordances est présenté à l'agent, et que cet état (cet ensemble d'affordances) fait partie du « mapping », alors le système perceptuel va poser l'action* que le « mapping » associe à cet état : le système perceptuel va décider de porter son attention sur l'une des affordances. Ainsi, si le « mapping » comprend le « mapping » ou l'association « "affordance x, y et z" ⇒ "décision de faire attention à l'affordance x" » et que l'état de l'environnement présente les affordances x, y et z, alors le système perceptuel (ou agent) dirigera son attention vers l'affordance x[70].

Et en ce qui concerne la quatrième notion, la notion de récompense, nous pouvons lui donner le sens suivant. À chaque fois que, grâce à son action*, l'agent cause un état de l'environnement et que cet état de l'environnement permet de réaliser l'intention de l'organisme, cet état de l'environnement donne une récompense (positive) à l'agent. Et nous pourrions associer une récompense négative à l'état de l'environnement qui rend impossible la réalisation de l'intention.

Je viens de poser que l'agent ou système perceptuel reçoit une récompense (positive) lorsque son action* (c.-à-d., plus précisément, une décision de faire attention à telle affordance plutôt qu'aux autres) cause un état qui permet de réaliser l'intention de l'organisme. On pourrait s'objecter contre cette possibilité; plus précisément, on pourrait faire valoir que l'action* de l'agent ne peut pas causer un état qui réalise l'intention de l'organisme. L'argument à la base de cette objection prend la forme suivante.

Les actions* de l'agent sont des décisions perceptuo-attentionnelles; plus précisément, l'action* de l'agent est une décision de faire attention à telle affordance plutôt qu'aux autres. Et les états de l'environnement qui permettent de réaliser une intention sont des états qui se

---

[70] J'ai dit ci-dessus que l'acte d'attention porté à une affordance n'était pas nécessairement un acte conscient et je dis maintenant que le processus attentionnel résulte d'une *décision* du système perceptuel à faire attention à l'affordance. Or, le terme « décision » réfère habituellement à quelque chose de conscient. Il faut donc être prudent ici et ne pas attribuer un caractère conscient à la décision du système perceptuel de porter attention à une affordance. Je suis conscient du caractère plutôt inapproprié du terme « décision », mais je suis ici fidèle à l'usage de Sutton et Barto (1998).

caractérisent par certaines caractéristiques physiques; par exemple, l'état environnemental qui réalise l'intention d'exécuter un coup droit se caractérise par une balle de tennis qui est renvoyée à l'adversaire à partir d'un coup de raquette qui a été exécuté du côté de la prise de raquette. Or, pourrait faire valoir notre objecteur, l'action* de l'agent, c'est-à-dire la décision de faire attention à telle affordance, ne cause pas des états de ce type. Cette objection ne tient pas; ces décisions causent effectivement des états de ce type. Comme on l'a dit, ces décisions font en sorte que le système perceptuel prête attention à une affordance. Or, au cours du chapitre 1, j'ai montré que l'attention portée à une affordance causait la perception d'une variable optique et que cette dernière perception causait un mouvement. Donc, par transitivité, la décision de l'agent de faire attention à telle affordance plutôt qu'aux autres cause un mouvement. Or, l'exécution de mouvements cause évidemment des états environnementaux du type qui vient d'être décrit, c'est-à-dire des états environnementaux qui satisfont des intentions. Donc, par transitivité, la décision de faire attention à une affordance cause un état environnemental qui peut satisfaire une intention.

Avant de définir la forme particulière que prend pour nous le problème de l'apprentissage par renforcement, il faut se confronter à une autre objection. Cette objection concerne la possibilité de baser nos actions* sur des politiques. Plus précisément, cette objection s'appuie sur le fait qu'une *même* situation (un même état environnemental) peut susciter des actions* différentes, c'est-à-dire des processus attentionnels différents. Par exemple, si je suis dans ma cuisine, qui me présente une multitude d'affordances, et que j'ai l'intention de me faire du café, mon système perceptuel fait attention à une affordance qui est différente de l'affordance à laquelle je fais attention si j'ai l'intention de me faire des pâtes. Dans un cas, je vais faire attention à l'affordance présentée par la cafetière; dans l'autre cas, à l'affordance présentée par le chaudron. En quoi le fait qu'une même situation puisse susciter des actions* différentes est-il problématique pour la thèse voulant que nos actions* soient basées sur une politique? Dans une politique, un état environnemental donné est associé à une action* donnée, et non pas à plusieurs actions*[71]. Donc, à première vue, on comprend mal comment un agent basant ses actions* sur une politique puisse poser des actions* différentes lorsqu'il est confronté à un même état environnemental.

---

[71] On pourrait répondre à cette objection en posant l'existence de politiques stochastiques. Pour l'instant, je ne tiens pas compte de cette réponse.

On peut aisément répondre à cette objection. Il suffit de poser l'existence de plusieurs politiques au sein du système perceptuel. Plus précisément, pour se soustraire à cette objection, il suffit de poser qu'il existe plusieurs politiques dont les « mappings » font référence au même état de l'environnement. Ainsi, lorsque le système perceptuel fait attention à des affordances différentes et qu'il est confronté à un même état environnemental, c'est qu'il est sous l'emprise de politiques différentes. Il reste à trouver l'élément qui détermine l'activation d'une politique plutôt qu'une autre. Il n'est pas difficile de trouver ce qui détermine l'activation d'une politique plutôt qu'une autre. Ce qui fait que je fais attention à cette affordance plutôt qu'à cette autre, c'est que j'ai une intention plutôt qu'une autre. Comme mon exemple précédent le montre, je fais attention à l'affordance présenté par la cafetière parce que j'ai l'intention de prendre du café et ce qui fait que je fais attention à l'affordance présentée par le chaudron, c'est que j'ai l'intention de me faire des pâtes. Donc, dans mon modèle, ce qui active une politique plutôt qu'une autre, c'est mon intention.

Précisons maintenant la forme que prend le problème de l'apprentissage par renforcement lorsque l'agent correspond à un système perceptuo-attentionnel et que les quatre notions (celles d'état environnemental, d'action*, de politique et de récompense) ont le sens qui vient de leur être donné. Le problème de l'apprentissage par renforcement prend pour nous la forme suivante : Comment, par son interaction avec l'environnement (les propriétés perceptibles de l'environnement, y compris les affordances), l'agent (le système perceptuo-attentionnel) peut-il changer sa politique actuelle en une politique optimale, c'est-à-dire une politique qui lui fournit un maximum de récompenses (en assignant de manière appropriée les récompenses, le maximum de récompenses correspond à la réalisation la plus efficace de l'intention)? Il est maintenant temps de présenter les méthodes qui fournissent une réponse à cette question.

8.2. La méthode de programmation dynamique[72]

La première méthode que je présente, la méthode de programmation dynamique (la méthode PD), n'est pas vraiment une méthode d'apprentissage par renforcement puisque la

[72] La programmation dynamique a fait l'objet de plusieurs ouvrages : Bertsekas (1995), Bertsekas et Tsitsiklis (1996), Dreyfus et Law (1977), Ross (1983), White (1969) et Whittle (1982, 1983).

politique optimale à laquelle elle conduit n'est pas le fruit d'interactions avec l'environnement. Pour trouver la politique optimale, cette méthode use, non pas d'interaction avec l'environnement, mais d'un modèle complet de l'environnement. Cette méthode est toutefois utile à connaître, car elle fournit une bonne introduction aux méthodes d'apprentissage par renforcement. Ma présentation de cette méthode se fait en deux étapes. Premièrement, je montre comment la méthode PD *évalue* la politique actuellement suivie par l'agent. Deuxièmement, je montre comment la méthode PD *améliore* cette politique au point de la rendre optimale.

8.2.1. L'évaluation de la politique actuelle

La première chose que la méthode PD accomplit, c'est d'évaluer la politique actuellement suivie par l'agent. Pour évaluer sa politique, l'agent a recours non pas à des expériences (des interactions) avec son environnement, mais à un modèle complet de la dynamique de l'environnement. Lorsque je dis que l'agent possède un modèle complet de la dynamique de l'environnement, je veux dire essentiellement deux choses. Premièrement, il connaît la probabilité des transitions d'états. Plus précisément, il sait que, s'il pose une action* donnée à partir d'un état donné de l'environnement, il aura telle probabilité de tomber dans tel autre état de l'environnement. Deuxièmement, il connaît la récompense immédiate qui lui est donnée lorsque, après avoir posé une action* à partir d'un état donné de l'environnement, il tombe dans un autre état de l'environnement[73].

Préalablement à la présentation de la méthode par laquelle l'agent évalue sa politique actuelle, il serait bon de savoir ce que l'agent cherche à connaître lorsqu'il évalue sa politique. Comme on l'a dit, le critère d'évaluation d'une politique est la quantité de récompenses qu'elle fournit. Plus une politique donne de récompenses, plus elle est bonne.

---

[73] La citation suivante atteste que, pour les problèmes d'apprentissage par renforcement, une connaissance complète du monde (plus précisément : une connaissance complète de la dynamique du monde) consiste à connaître les probabilités des transitions d'état et les récompenses attendues : « A reinforcement learning problem can be posed in a variety of different ways depending on assumptions about the level of knowledge initially available to the agent. In problems of *complete knowledge*, the agent has a complete and accurate model of the environment's dynamics. (...)(S)uch a model consists of the one-step *transition probabilities* and *expected rewards* for all states and their allowable actions. In problems of *incomplete knowledge*, a complete and perfect model of the environment is not available. » (Sutton et Barto, 1998, p. 82)

Donc, ce que l'agent cherche à connaître lorsqu'il évalue sa politique, c'est de savoir combien elle donne de récompenses. L'agent peut savoir cela de deux façons : 1) en évaluant les états de l'environnement, c'est-à-dire en déterminant, pour chaque état de l'environnement, les récompenses (immédiates et éloignées) qui font suite à la visite de l'état lorsque l'agent suit sa politique (la politique qu'il veut évaluer); ou 2) en évaluant les paires état-action*, c'est-à-dire en déterminant, pour chaque action* exécutée à partir d'un état, les récompenses (immédiates et éloignées) qui font suite à l'action* exécuté à partir de l'état. C'est en procédant à l'un ou l'autre de ces types d'évaluations que l'agent évalue une politique. Mais comment évalue-t-on un état et une paire état-action*?

Pour présenter la méthode PD par laquelle un état est évalué, il faut d'abord préciser ce qu'il faut comprendre par la valeur d'un état. La valeur d'un état correspond aux récompenses auxquelles nous sommes en droit de nous attendre lorsque l'agent part de cet état et qu'il suit par la suite la politique qui doit être évaluée, disons la politique π. Il est important de souligner que les récompenses qui suivent la visite de l'état s[74] ne correspondent pas seulement aux récompenses immédiates, celles qui surviennent lorsque l'agent « tombe » dans l'état qui suit immédiatement l'état s, c'est-à-dire lorsqu'il « tombe » dans l'état s'. Les récompenses dont il est question comprennent, en plus des récompenses immédiates (les récompenses données par s'), les récompenses qui viennent après avoir visité s', c'est-à-dire les récompenses qui viennent au contact des états s'', s''', etc. L'ensemble de ces récompenses constitue le *retour* de l'état s. On peut donc dire que la valeur de l'état s correspond au retour auquel l'agent est en droit de s'attendre après avoir visité l'état *s*, cela s'il suit la politique π. Formellement, cette définition de la valeur de l'état s peut être décrite de la manière suivante : $V^{\pi}(s) = E_{\pi}(R_t | s_t = s)$. La valeur d'un état peut être définie d'une autre manière, équivalente à celle qui vient d'être présentée. Cette façon équivalente de définir la valeur d'un état est exprimable par la description symbolique suivante :

$$V^{\pi}(s) = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^{a} [\Re_{ss'}^{a} + \gamma V^{\pi}(s')]$$[75]

---

[74] La variable « s » est mise pour le mot « state » ou, en français, « état ». Je n'ai pas utilisé la variable « e » pour être conforme au formalisme de Sutton et Barto, qui sera présenté plus loin.

[75] Pour une démonstration mathématique de l'équivalence entre ces deux équations, voir Sutton et Barto, 1998, p. 70.

Je vais maintenant tenter de traduire en langage ordinaire ce que cette équation nous dit en si peu de symboles. Cette équation nous dit d'abord que pour trouver la valeur de l'état s, il faut considérer la récompense qui suit immédiatement la visite de l'état s et l'action* a, c'est-à-dire la récompense qui survient lorsque l'agent tombe dans l'état s' après avoir exécuté l'action* a à partir de l'état s, et qu'il faut additionner cette récompense immédiate à la *valeur* de l'état s', l'état suivant. Cette addition correspond à l'expression symbolique « $[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')]$ ». Ensuite, il faut multiplier cette somme à une probabilité. En effet, le fait que l'agent tombe dans l'état s' après avoir exécuté l'action* a à partir de l'état s peut être simplement probable. L'agent aurait pu exécuter la même action* à partir du même état et il aurait pu tomber dans un autre état. Il faut donc pondérer cette somme (récompense immédiate + valeurs de l'état s') à la probabilité de tomber dans l'état s' après avoir exécutée l'action* a à partir de l'état s. Autrement dit, il faut multiplier la somme « $(\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s'))$ » à la probabilité $P_{s}^{a}{}_{s'}$; il faut calculer : $P_{s}^{a}{}_{s'}[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')]$. Et s'il y a effectivement une probabilité pour qu'un autre état s' suive l'action* a exécutée à partir de l'état s, il faut faire une deuxième fois le calcul qui vient d'être décrit, cette fois pour cet autre transition d'état. Et on additionne ensuite ces deux calculs. Et s'il s'avère qu'il existe une troisième transition d'état possible, voire une $n^{ième}$ transition d'état, il faut faire un calcul pour chacune de ces transitions d'état, et il faut ensuite les additionner toutes. Cette somme totale $(\sum_{s'} P_{s}^{a}{}_{s'}[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')])$ doit ensuite être pondérée à la probabilité $\pi(s, a)$, c'est-à-dire à la probabilité que, suivant la politique π, l'agent exécute l'action* a à partir de l'état s. En effet, la politique π peut être stochastique. Ce qui vient d'être dit correspond à cette portion de l'équation : $\pi(s, a)\sum_{s'} P_{s}^{a}{}_{s'}[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')]$. Et, si la politique π est effectivement stochastique, il faut répéter le même calcul pour les autres actions* probables que la politique π prescrit. Tous les calculs « $\pi(s, a)\sum_{s'} P_{s}^{a}{}_{s'}[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')]$ » doivent ensuite être additionnés. La somme correspond à la valeur de l'état s, $V^{\pi}(s)$.

La valeur d'une paire état-action*, $Q^{\pi}(s, a)$, peut être définie de la manière suivante. C'est le retour auquel l'agent est en droit de s'attendre lorsqu'il part de l'état s et qu'il exécute l'action* a, tout en suivant par la suite la politique π. Une définition équivalente à cette définition existe. Formellement, cette définition équivalente est celle-ci : $Q^{\pi}(s, a) = \sum_{s'} P_{s}^{a}{}_{s'}[\Re_{s}^{a}{}_{s'} + \gamma V^{\pi}(s')]$. Cette équation est identique à celle qui donne la valeur de l'état s,

sauf que, dans l'équation présente, il n'est pas question de faire la somme des actions* probables, $\sum_a \pi(s, a)$. Cela est compréhensible : l'évaluation d'une paire état-action* (s, a) ne cherche pas à connaître la probabilité des autres actions* qui peuvent être exécutées à partir de l'état s; elle cherche à savoir ce qui se passe lorsque l'action* a est effectivement exécutée à partir de l'état s. Donc, dans cette perspective, les autres actions* qui peuvent être exécutées à partir de l'état s ne sont pas considérées et la probabilité $\pi(s, a)$ que l'action* a soit exécutée à partir de l'état s est de 1. On comprend donc pourquoi la somme $\sum_a \pi(s, a)$ a été éliminée de la formule permettant de calculer la valeur de la paire état-action* (s, a), $Q^{\pi}(s, a)$, et que, à la place, on ait introduit la quantité 1.

Maintenant que nous savons de quoi il en retourne lorsqu'il est question de la valeur d'un état et d'une paire état-action*, il faut s'interroger sur la façon de trouver la valeur d'un état et d'une paire état-action*. En fait, ici, suivant Sutton et Barto (1998), je vais me contenter de présenter une méthode d'évaluation pour les états, et non pas pour les paires état-action*. La méthode d'évaluation que je vais d'abord présenter ne relève pas de la méthode PD. Je prends le temps de la présenter car elle offre une bonne entrée en matière.

La plupart du temps, les retours reçus après la visite d'un état donné ne sont pas identiques. La variabilité peut venir de plusieurs facteurs. D'une part, la politique de l'agent peut être stochastique, si bien que, lorsqu'il part d'un état donné, l'agent peut poser des actions* différentes de fois en fois, et cela vaut aussi pour tous les états qu'il va visiter par la suite. Cela fait évidemment en sorte que le retour peut varier de fois en fois. D'autre part, même si la politique était déterministe, les transitions d'état peuvent être seulement probables. Ainsi, à partir d'un état donné et d'une action* donnée, la probabilité que le prochain état soit x peut être de β et qu'il soit y peut être de $1-\beta$.

Cependant, malgré cette variabilité, l'agent peut trouver la valeur réelle d'un état. Si l'agent passe par un état un très grand nombre de fois, la valeur *moyenne* des retours devrait converger vers la valeur réelle de l'état. C'est donc en faisant la moyenne d'un très grand nombre de retours que la valeur d'un état est trouvée. Cette méthode d'évaluation — faire la moyenne d'un très grand nombre de retours — est la méthode utilisée par la seconde méthode d'apprentissage par renforcement, la méthode Monte-Carlo, qui sera présentée plus tard. Abordons maintenant la méthode d'évaluation PD.

La méthode d'évaluation PD calcule par itération la valeur de l'état s. En fait, il s'agit de répéter un très grand nombre de fois un calcul qui est essentiellement similaire au membre de droite de l'équation présentée plus haut. La valeur de l'état s après une itération, $V_{k+1}(s)$, est donnée par l'équation suivante : $V_{k+1}(s) = \sum_a \pi(s, a) \sum_{s'} P^a_{ss'} [\Re^a_{ss'} + \gamma V_k(s')]$. Il faut répéter ce calcul un très grand nombre de fois sur chaque état de l'environnement, et une convergence vers la valeur réelle de ces états devrait être observée. Au tout début, il faut initialiser la valeur des états à 0. Aussi, il faut noter qu'avant de réitérer ce calcul sur un état donné de l'environnement, il faut s'assurer d'avoir appliqué ce calcul aux autres états de l'environnement. Il faut donc « balayer » ce calcul à l'ensemble des états de l'environnement, et une fois que le « balayage » est complété, il faut réitérer ce « balayage ». Pour savoir quand ce « balayage » itératif prend fin, il faut calculer la différence entre la dernière itération et l'avant-dernière itération : lorsque cette différence est très petite, voire pratiquement nulle, la valeur réelle des états est trouvée. C'est ainsi que la méthode PD procède pour trouver la valeur $V^\pi(s)$ de chaque état s. La figure 8.1 est un exemple de la méthode itérative d'évaluation.

Dans cet exemple, il faut observer les tableaux de la colonne de gauche. Plus les tableaux sont situés vers le bas, plus l'itération du « balayage » a été effectuée un grand nombre de fois. Les états de l'environnement correspondent aux cases des tableaux. La politique suivie par l'agent est « à l'aveuglette », c'est-à-dire que, lorsque l'agent est dans une case du tableau, ses chances de poser l'action* « vers le haut », « vers le bas », « vers la droite » ou « vers la gauche » sont égales. $\pi(s, a)$ est donc de 0.25. Les deux états finaux sont la case en bas à droite et la case en haut à gauche. Toutes les cases donnent une récompense de -1. $\Re^a_{ss'}$ a donc une valeur de -1 pour tous les états. Les transitions d'état s'effectuent avec une probabilité de 1. Autrement dit, si l'agent est dans une case (un état) donnée et qu'il exécute une action* donnée, la probabilité pour qu'il se transporte dans l'autre état est de 1. $P^a_{ss'}$ prend donc la valeur de 1. Si l'agent est dans une case qui est au bord du tableau et qu'il effectue une action* vers l'extérieur du tableau, l'agent reste sur sa case avec une probabilité de 1. Avec ces informations, on peut « balayer » itérativement le calcul $\sum_a \pi(s, a) \sum_{s'} P^a_{ss'} [\Re^a_{ss'} + \gamma V_k(s')]$ sur tous les états et on devrait arriver aux valeurs réelles des états, c'est-à-dire à V^π.

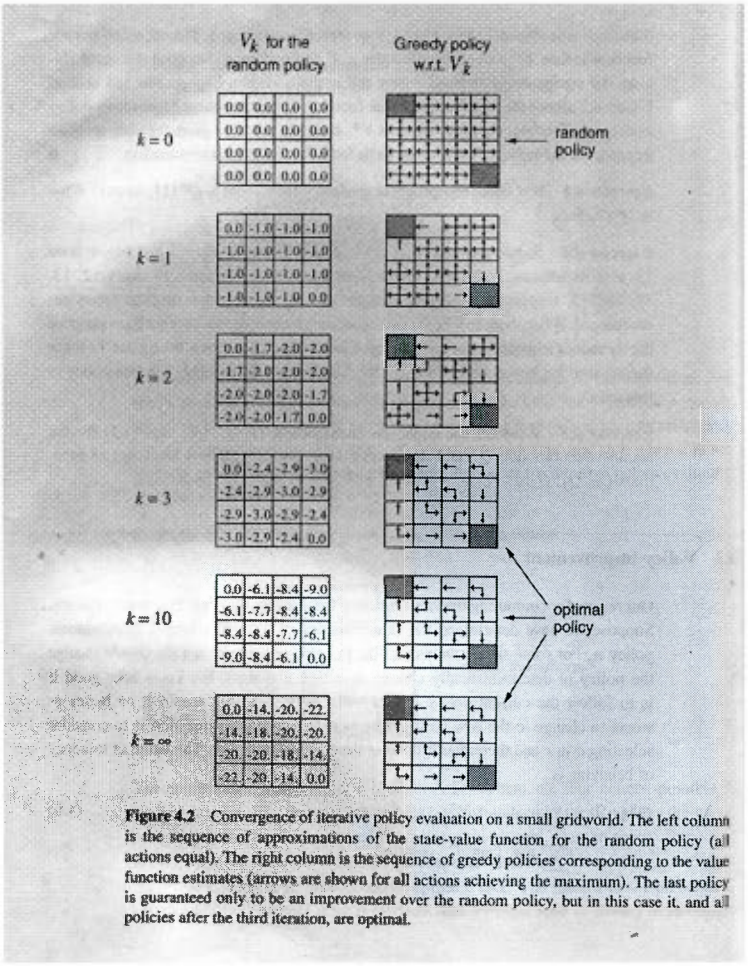


Figure 4.2 Convergence of iterative policy evaluation on a small gridworld. The left column is the sequence of approximations of the state-value function for the random policy (all actions equal). The right column is the sequence of greedy policies corresponding to the value function estimates (arrows are shown for all actions achieving the maximum). The last policy is guaranteed only to be an improvement over the random policy, but in this case it, and all policies after the third iteration, are optimal.

Figure 8.1 Exemple de la méthode itérative d'évaluation. Tirée de Sutton et Barto (1998).

Mais est-ce que l'évaluation des états équivaut réellement à une évaluation de la politique? Absolument. La politique π est évaluée à partir du moment où les récompenses qu'elle rapporte sont connues; or, l'évaluation des états de l'environnement, V^{π}, nous dit précisément combien de récompenses la politique π rapporte.

Il faut prêter attention à une dernière chose avant de passer à la prochaine section. La formule itérative permettant de calculer la valeur d'un état — $V_{k+1}(s) = \sum_a \pi(s, a) \sum_{s'} P^a_{ss'} [\Re^a_{ss'} + \gamma V_k(s')]$ — requiert que l'agent connaisse la probabilité $\pi(s, a)$ d'exécuter l'action* a à partir de l'état s. La connaissance de cette probabilité ne pose pas de problème pour l'agent puisque cette probabilité relève de l'agent lui-même. Mais, pour faire ce calcul, et c'est ce qui est important de souligner, l'agent a aussi besoin de connaître, d'une part, la probabilité $P^a_{ss'}$,

de tomber dans l'état s' après avoir exécuté l'action* a à partir de l'état s et, d'autre part, de connaître la récompense $\Re_{s}^{a}{}_{s'}$ qui est donnée à l'agent lorsqu'il tombe dans l'état s' après avoir exécuté l'action* a à partir de l'état s[76]. Conformément à ce qui a été dit plus haut, la connaissance de ces deux quantités équivaut à une connaissance du monde. Donc, comme on l'a dit, l'évaluation PD de la politique exige que l'agent ait en sa possession un modèle complet du monde. Comme je l'ai dit plus haut, c'est l'un des points importants sur lequel la méthode PD se démarque des méthodes d'apprentissage par renforcement. Les méthodes d'apprentissage par renforcement n'évaluent pas leurs politiques à partir d'un modèle du monde, mais à partir d'interactions avec le monde.

Maintenant que j'ai montré comment évaluer la politique actuellement suivie par l'agent, nous allons voir comment la politique actuelle peut être améliorée, jusqu'au point de la rendre optimale.

8.2.2. Vers une politique optimale

Avant de montrer comment l'agent parvient à trouver une politique optimale, il faut d'abord montrer comment il parvient à améliorer sa politique actuelle.

8.2.2.1. L'amélioration de la politique

On améliore une politique π en rendant l'agent « glouton » (greedy) relativement à l'évaluation que cet agent a faite des états de son environnement, c'est-à-dire « glouton » relativement à V^{π}. Je m'explique.

À partir de l'état s, l'agent peut faire une multitude d'actions*. Et chacune de ces actions* mène à un (ou plusieurs) état(s) s' différent(s). L'agent sait quels sont ces états. En effet, notre agent possède un modèle complet du monde. Plus précisément, notre agent possède une connaissance de toutes les probabilités de transitions d'état ($P_{s}^{a}{}_{s'}$). De plus, nous connaissons la valeur de chacun de ces états s'. C'est en effet au cours de l'évaluation des états que l'agent a pris connaissance de leur valeur. Maintenant, qu'est-ce que nous voulons dire par :

---

[76] La formule itérative fait aussi référence à la valeur $V_k(s')$ de l'état s'. Comme on l'a dit, initialement, la valeur $V_k(s')$ de l'état s' est arbitrairement donnée. Et à mesure qu'il y a itération de la formule itérative, la valeur $V_k(s')$ de l'état s' change, cela jusqu'au moment où cet état atteint sa valeur réelle.

« glouton » relativement à V^{π}? Nous voulons simplement dire que l'agent intègre dans sa nouvelle politique l'action* qui mène à l'état s' qui, si la récompense immédiate donnée par cet état est aussi prise en compte, possède la plus grande valeur. Il faut procéder de cette façon « gloutonne » à partir de chaque état. À l'issue de ce processus, l'agent tient en main une nouvelle politique (la politique π') qui donne plus de récompenses que la politique π et qui, donc, représente une amélioration relativement à la politique π. Si la politique π' n'est pas meilleure que la politique π, c'est que la politique π était déjà une politique optimale. « Policy improvement thus must give us a strictly better policy except when the original policy is already optimal. » (Sutton et Barto, 1998, p. 96)

8.2.2.2 La politique optimale

Après avoir amélioré notre politique, il faut évaluer cette nouvelle politique conformément à la méthode d'évaluation présentée plus haut. Il faut donc évaluer $V^{\pi'}$. Ensuite, il faut améliorer cette nouvelle politique en rendant l'agent « glouton » relativement à $V^{\pi'}$. À l'issue de ce processus, une nouvelle politique, la politique π'', est formée, et celle-ci est une amélioration relativement à la politique π'. On répète ensuite sur la politique π'' les opérations qui ont été appliquées aux politiques π et π' : on évalue la politique π'' et on l'améliore en rendant l'agent « glouton ». En continuant ce processus itératif, on devrait converger vers la politique optimale. La politique optimale est trouvée lorsque la dernière politique améliorée n'est pas différente de la précédente.

8.3. La méthode Monte-Carlo[77]

Pour trouver la politique optimale, la méthode PD se sert d'un modèle complet du monde, et non pas des expériences de l'agent. À ce titre, comme nous l'avons déjà souligné, il ne s'agit pas à proprement parler d'une méthode d'apprentissage par renforcement, puisque l'apprentissage par renforcement procède sur la base d'interactions avec le monde. Mais, avant de présenter les méthodes d'apprentissage par renforcement proprement dites, il faut se demander si l'apprentissage du système perceptuo-attentionnel est conforme à la méthode PD ou, au contraire, s'il est conforme aux méthodes d'apprentissage par renforcement. D'une

[77] Voir par exemple : Kalos et Whitlock (1986) et Rubinstein (1981).

part, il me semble clair que l'apprentissage qui est à la base de notre capacité à faire attention aux bonnes affordances est un apprentissage qui requiert des interactions avec l'environnement. D'autre part, le simple fait de posséder un modèle complet du monde n'est pas une thèse psychologique très plausible. Ces deux constats m'incitent à présenter de vraies méthodes d'apprentissage par renforcement. Je vais commencer par présenter la méthode d'apprentissage par renforcement nommée « méthode Monte-Carlo ».

Les étapes que les méthodes d'apprentissage par renforcement doivent suivre pour trouver la politique optimale se conforment aux étapes de la méthode PD. Il s'agit d'abord d'évaluer la politique en évaluant les états ou les paires état-action*. Il s'agit ensuite d'améliorer la politique en rendant l'agent « glouton » relativement à la valeur des états ou des paires état-action*. Ensuite, il faut répéter ces deux étapes sur la nouvelle politique, améliorée. En fait, il faut réappliquer ces deux étapes sur chaque nouvelle politique améliorée. En procédant de cette manière, c'est-à-dire itérativement, l'agent devrait converger vers la politique optimale.

En grande partie, ce qui distingue les méthodes d'apprentissage entre elles, c'est la première étape, l'évaluation de la politique (qui se fait à travers l'évaluation des états ou des paires état-action*). Ce qui distingue principalement la méthode d'évaluation Monte-Carlo de la méthode d'évaluation PD, c'est que la méthode Monte-Carlo évalue sa politique (évalue les états et les paires état-action*), non pas sur la base d'un modèle complet de l'environnement, mais sur la base d'expériences avec cet environnement. Je vais maintenant montrer comment la méthode Monte-Carlo procède pour évaluer la politique actuelle de l'agent.

8.3.1. L'évaluation Monte-Carlo de la politique

Pour évaluer un état, l'agent doit d'abord faire l'expérience de cet état, que ce soit au début de l'épisode ou en cours d'épisode. Ensuite, l'agent tient compte des récompenses qu'il reçoit après avoir visité cet état. Plus précisément, il calcule les récompenses qui font suite à la visite de l'état, cela jusqu'au moment où l'épisode prend fin ou jusqu'au moment où il refait l'expérience du même état[78]. Autrement dit, l'agent calcule le retour. Pour les raisons

[78] Il existe une méthode d'évaluation Monte-Carlo qui tient compte des récompenses qui font suite à un deuxième (ou un $n^{ième}$) passage sur l'état pour lequel l'agent calcule les récompenses. Autrement

que nous avons présentées plus haut, le retour d'un état n'est pas toujours identique. On ne peut donc pas se fier à une seule expérience pour connaître la valeur réelle d'un état. Le moyen utilisé par la méthode Monte-Carlo pour trouver la valeur de l'état est bien simple. Elle fait la moyenne des retours. Selon la loi des grands nombres, si cette moyenne est basée sur un très grand nombre de retours, la moyenne est censée converger vers la valeur réelle de l'état. Donc, avant de connaître la valeur des états de l'environnement, l'agent doit faire une multitude d'expériences. En fait, ce nombre d'expériences est plus ou moins grand selon que la politique est plus ou moins stochastique ou que les transitions d'état produites par une action* exécutée à partir d'un état donné sont plus ou moins probables. À l'un des extrêmes, si la politique est très stochastique et que, sur la base d'une action* exécutée à partir d'un état donné, il existe une multitude de transitions d'états probables, l'agent aura besoin de beaucoup plus de retours pour établir une moyenne qui corresponde à la valeur réelle de l'état. Et, à l'autre extrême, si la politique est entièrement déterministe et que les transitions d'états possèdent toutes une probabilité de 1, alors la connaissance de la valeur réelle des états ne requiert qu'un seul retour.

Un point important : Pour évaluer sa politique, la méthode Monte-Carlo n'évalue pas les états, mais les paires état-action*. C'est que, à défaut de la valeur des paires état-action*, l'agent de la méthode Monte-Carlo ne peut pas *améliorer* sa politique. L'agent de la méthode Monte-Carlo ne possède pas un modèle complet de l'environnement; il ne possède pas les probabilités des transitions d'état. Cet agent ne peut donc pas se projeter dans un état situé plus loin. Or, lorsque l'agent améliore sa politique et que cette amélioration s'appuie sur la valeur des états, comme ce fut le cas pour la méthode PD présentée plus haut, il faut précisément que l'agent aille voir la valeur des états qui sont situés une étape plus loin, c'est-à-dire la valeur des états s'. Donc, si l'agent de la méthode Monte-Carlo ne disposait que de la valeur des états, il ne pourrait pas améliorer sa politique, faute de posséder un modèle complet du monde et, donc, de pouvoir se projeter une étape plus loin. Ce problème, typique de l'agent qui ne possède pas un modèle complet du monde, ne se pose pas lorsque l'amélioration de la politique s'appuie sur la valeur des paires état-action*. En effet, pour connaître la valeur d'une action* exécutée à partir d'un état s, c'est-à-dire la valeur d'une

---

dit, selon cette autre méthode d'évaluation, seul le terme de l'action met terme au calcul des récompenses faisant suite à la visite de l'état.

paire état-action*, l'agent n'a pas à se projeter un état plus loin; l'action* réfère à l'état présent; elle est contemporaine de l'état actuel, l'état s.

Donc, comme l'amélioration de la politique s'effectue à partir de la valeur des paires état-action*, il est important de savoir comment évaluer les paires état-action*.

La méthode d'évaluation des paires état-action* ne diffère pas beaucoup de la méthode d'évaluation des états. Pour connaître la valeur de la paire état-action* (s, a), il faut exécuter l'action* a à partir de l'état s, suivre ensuite la politique π et récolter toutes les récompenses qui suivent; autrement dit, il faut recevoir le retour qui suit l'action* a (exécutée à partir de l'état s). L'agent doit répéter souvent l'action* a (exécutée à partir de l'état s) afin de recevoir plusieurs retours. L'agent fait la moyenne de ces retours, et si le nombre de retour est suffisamment grand, la valeur réelle de l'action* a, exécutée à partir de l'état s, devrait correspondre à cette moyenne.

La méthode d'évaluation Monte-Carlo qui a été présenté est lourde computationnellement. Pour chaque moyenne que l'agent veut établir, il doit se souvenir de tous les retours passés, il doit les additionner et, enfin, il doit diviser cette somme par le nombre de retours. À la place de cette méthode de calcul, il est possible d'adopter une méthode de calcul incrémentielle. L'assignation permettant d'évaluer un état et l'assignation permettant d'évaluer une paire état-action* sont formellement identiques.

Pour évaluer un état s, l'assignation est : $V(s_t) \leftarrow V(s_t) + \alpha(R_t - V(s_t))$. Et pour évaluer une paire état-action*, il s'agit de l'assignation suivante : $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(R_t - Q(s_t, a_t))$. La forme générale de ces deux assignations est la suivante : *NouvelleÉvaluation* $\leftarrow$ *VieilleÉvaluation* + *StepSize(Retour-VieilleÉvaluation)*. Il s'agit donc d'abord de soustraire l'ancienne valeur de l'état (ou de la paire état-action*) au retour. On additionne ensuite cette différence à l'ancienne valeur de l'état (ou de la paire état-action*). Cette somme est la nouvelle valeur de l'état (ou de la paire état-action*). Et l'évaluation de l'état (ou de la paire état-action*) a toutes les chances de correspondre à la valeur réelle de l'état (ou de la paire état-action*) si l'agent fait l'expérience d'un très grand nombre de retours et qu'il applique donc cette formule un très grand nombre de fois.

Pour terminer, il faudrait dire quelques mots sur le paramètre « step-size ». Celui-ci influence le taux d'apprentissage, et, dans la plupart des cas, il correspond à une fraction qui diminue de fois en fois, c'est-à-dire qui diminue à chaque fois qu'un nouveau retour se présente et que la formule en vue de trouver la valeur de l'état (ou de la paire état-action*) est appliquée. La diminution du paramètre « step-size » peut être justifiée. Plus le nombre de retours est grand, plus on doit converger vers la valeur réelle de l'état (ou de la paire état-action*). Donc, plus le nombre de retours précédant le retour actuel est grand, moins la différence qui existe entre le retour actuel et l'ancienne évaluation (Retour-VieilleEvaluation) doit pouvoir changer la nouvelle valeur de l'état (ou de la paire état-action*). On parvient à ce résultat en donnant au paramètre « step-size » une valeur fractionnaire de plus en plus petite à mesure que les retours s'accumulent.

8.3.2 Le problème lié à l'amélioration de la politique et la solution à ce problème

J'aimerais maintenant présenter un problème qui touche à la fois la méthode Monte-Carlo et la méthode des différences temporelles (la méthode d'apprentissage par renforcement qui sera bientôt présentée). La solution que Sutton et Barto (1998) apportent à ce problème vaut autant pour la méthode Monte-Carlo que pour la méthode des différences temporelles. Il faudra donc garder en tête cette solution lorsqu'il sera question de la méthode des différences temporelles.

Ce problème concerne l'amélioration de la politique (c.-à-d. la deuxième étape du processus itératif menant à la politique optimale) et il apparaît uniquement lorsque la politique suivie par l'agent est une politique déterministe.

Si l'agent possède une politique déterministe, alors, pour chaque état, il fera l'expérience d'une seule action*. Autrement dit, il n'aura jamais deux paires état-action* qui réfèrent au même état. Donc, si la politique est déterministe, l'évaluation des paires état-action* ne présentera jamais à l'agent un choix de paires état-action*. Or, l'amélioration de la politique suppose que l'agent soit confronté à au moins deux paires état-action*, puisqu'il doit choisir la meilleure paire état-action*. Une politique déterministe semble donc rendre impossible l'amélioration de la politique. Pour remédier à ce problème, il faut que l'agent fasse des actions* *exploratoires*.

Pour donner une dimension exploratoire à l'agent, Sutton et Barto (1998) suggèrent l'existence de débuts exploratoires. L'existence de débuts exploratoires signifie ceci : l'agent débute son épisode avec une paire état-action* (autrement dit, il débute l'épisode en exécutant une action* donnée à partir d'un état donné) et chaque paire état-action* possède une probabilité non nulle de commencer un épisode. Et après son commencement exploratoire, l'agent suit sa politique *déterministe*. De cette manière, toutes les paires état-action* seront essayées un très grand nombre de fois et il sera possible d'établir leur valeur. L'agent pourra donc choisir, parmi les paires état-action* qui font référence au même état, celle qui donne le plus de récompenses ou, autrement dit, celle qui possède la plus grande valeur.

L'hypothèse des débuts exploratoires n'est pas très plausible. L'hypothèse des débuts exploratoires pose que toutes les paires état-action* ont la possibilité de commencer un épisode. Dans nombre de cas, cela ne semble pas plausible. Dans de nombreuses situations d'action, certains états ne se retrouvent jamais au commencement de l'épisode; ceux-ci n'apparaissent qu'une fois que l'action (l'épisode) est bien entamée. Donc, la valeur des paires état-action* qui font référence à ce genre d'état ne pourra jamais être connue. Pour ces états, l'agent sera donc contraint de sélectionner la paire état-action* qui est donnée dans la politique déterministe; il n'aura pas d'autres choix.

Sutton et Barto (1998) présente une autre solution au problème lié aux politiques déterministes. Cette autre solution consiste à nier que l'apprentissage puisse se faire lorsque l'agent adopte une politique déterministe et à affirmer au contraire que, lors de l'apprentissage, la politique de l'agent est toujours stochastique. De cette façon, il est certain que toutes les paires état-action* seront essayées un très grand nombre de fois et que, donc, la valeur de chaque paire état-action* sera accessible à l'agent. Et s'il faut que la politique *optimale* soit une politique déterministe, il est possible de changer le caractère stochastique de la politique au cours de l'apprentissage. On peut rendre la politique de plus en plus déterministe au cours de l'apprentissage, quitte à lui donner un caractère entièrement déterministe au terme de l'apprentissage.

8.4. La méthode des différences temporelles[79]

Pour trouver la politique optimale, la méthode des différences temporelles (la méthode DT) suit les étapes générales des méthodes PD et Monte-Carlo: il s'agit d'évaluer la politique actuellement suivie, d'améliorer cette politique en rendant l'agent « glouton » relativement à la valeur des paires état-action* et de répéter ces deux étapes sur chaque politique améliorée. C'est en suivant ce processus itératif que l'agent devrait converger vers la politique optimale.

Encore une fois, ce qui distingue les trois méthodes, c'est la façon d'évaluer la politique actuelle de l'agent. Essayons donc de décrire la méthode DT d'évaluation.

La méthode DT d'évaluation se fonde aussi sur des expériences pour évaluer les états ou les paires état-action*. L'évaluation DT des états est plus intuitive que l'évaluation DT des paires état-action*. Je présenterai donc d'abord l'évaluation DT des états.

Comme la méthode Monte-Carlo, la méthode DT s'appuie sur un très grand nombre d'expériences pour évaluer un état, mais, contrairement à la méthode Monte-Carlo, l'agent n'attend pas la fin de l'épisode pour mettre à jour la valeur de l'état. Autrement dit, la méthode DT ne se sert pas des retours pour évaluer l'état. Pour évaluer l'état s, la méthode DT d'évaluation se sert de la récompense immédiate qui est donnée par l'état suivant, s', ainsi que de la valeur de ce dernier état. Ainsi, il faut remplacer la variable R (le retour) de l'assignation « $V(s) \leftarrow V(s) + \alpha(R - V(s))$ » par l'expression « $r + \gamma V(s')$ ». La méthode DT calcule donc V(s) en utilisant l'assignation : $V(s) \leftarrow V(s) + \alpha(r + \gamma V(s') - V(s))$.

Plus précisément, voici comment il faut procéder pour trouver la valeur des états. Il faut d'abord initialiser arbitrairement la valeur des états de l'environnement. Ensuite, l'agent débute un épisode. Supposons que l'épisode débute avec l'état s. Pour mettre à jour la valeur de l'état s, l'agent additionne d'abord la récompense qui lui est donnée lorsqu'il tombe dans l'état s' et la valeur qui, conformément à l'initialisation, a été donnée à l'état s'. Il soustrait ensuite la valeur qui a été initialement donnée à l'état s à cette somme. Il additionne ensuite cette différence à la valeur initiale de l'état s. Cette somme constitue alors la nouvelle valeur de l'état s. Tout ce que je viens de dire se résume à l'assignation présentée plus haut : $V(s) \leftarrow V(s) + \alpha(r + \gamma V(s') - V(s))$. Après avoir mis à jour la valeur de l'état s, il s'agit de mettre à

[79] Sutton (1988, 1996) a traité en profondeur de la méthode des différences temporelles.

jour la valeur de l'état s'. Il faut alors faire pour l'état s' ce qui a été fait pour l'état s. Après avoir procédé de la sorte pour l'état s', il faut répéter l'opération pour l'état s'' et pour tous les autres états de l'épisode, cela jusqu'à l'état terminal. Lorsque l'épisode est terminé, on répète tous ce processus en débutant un nouvel épisode et en tenant compte des nouvelles valeurs qui ont été données au cours de l'épisode précédent. Après la répétition d'un très grand nombre d'épisodes, la valeur de chaque état est censée converger vers sa valeur réelle. Grosso modo, ce qui précède décrit l'algorithme permettant de trouver la valeur des états.

Maintenant que la façon d'évaluer l'état s a été présentée, la façon d'évaluer la paire état-action* (s, a) peut être abordée. Et il faut noter que, pour trouver la politique optimale, il est nécessaire de trouver la valeur, non pas des états, mais des paires état-action*. Comme pour l'agent de la méthode Monte-Carlo, l'agent de la méthode DT ne dispose pas d'un modèle complet du monde, si bien qu'il ne peut pas se projeter une étape plus loin et que, à ce titre, il a besoin de connaître la valeur des paires état-action* pour améliorer sa politique.

L'assignation permettant de mettre à jour la valeur d'une paire état-action* prend la même forme que l'assignation permettant de mettre à jour la valeur de l'état s, sauf que, dans cette nouvelle assignation, il faut remplacer la valeur des états par la valeur de paires état-action*. La forme de cette nouvelle assignation est la suivante :

$$Q(s, a) \leftarrow Q(s\ a) + \alpha(r + \gamma Q(s', a') - Q(s, a))$$

Cette assignation nous dit qu'il faut tenir compte de la récompense immédiate qui est donnée par l'état s', après que l'agent ait exécuté l'action* a à partir de l'état s, et qu'il faut additionner cette récompense immédiate à la valeur de la paire état-action* (s', a'). Ensuite, il faut soustraire l'ancienne valeur de la paire action* (s, a) à cette somme. Cette différence doit ensuite être additionnée à l'ancienne valeur de l'état-action* (s, a). Cette dernière somme correspond à la nouvelle valeur de la paire état-action* (s, a).

L'algorithme permettant de trouver la valeur de l'ensemble des paires état-action* est similaire à l'algorithme permettant de trouver la valeur des états. Il s'agit d'abord d'initialiser arbitrairement l'ensemble des paires état-action*. L'agent débute ensuite un épisode. L'agent fait alors l'expérience d'une première paire état-action*; autrement dit, il exécute une action* à partir du premier état rencontré dans cet épisode. La valeur de cette première paire état-

action*, (s, a), est alors mis à jour conformément à l'assignation présentée dans le paragraphe précédent. La valeur de la seconde paire état-action*, (s', a'), est ensuite à son tour mise à jour à l'aide de la même assignation. Il faut répéter ce processus pour chaque paire état-action* rencontrée au cours de l'épisode, cela jusqu'au moment où l'agent atteint une paire état-action* terminale. Il faut ensuite commencer un nouvel épisode, en appliquant la même assignation pour chaque paire état-action* rencontrée en cours d'épisode. On répète le même processus sur plusieurs épisodes, et cela devrait nous permettre de converger vers la vraie valeur des paires état-action*.

Rappelons aussi que, pour que l'agent ait le choix entre plusieurs paires état-action* au moment d'améliorer sa politique, l'évaluation des paires état-action* doit s'effectuer sur la base d'une politique stochastique.

Comme je l'ai dit à maintes reprises, les étapes qui succèdent à l'évaluation de la politique et qui permettent de trouver la politique optimale sont similaires à la méthode PD et la méthode Monte-Carlo. Après avoir évalué sa politique, l'agent sélectionne « gloutonnement » les actions* qu'il décide de mettre dans sa nouvelle politique, et cela devrait faire de cette nouvelle politique une politique améliorée. Il faut ensuite répéter ces deux étapes sur cette nouvelle politique. En répétant l'application de ces deux étapes sur chaque nouvelle politique, on devrait converger vers la politique optimale.

Quand l'action se déroule au sein d'épisodes relativement longs, je considère que la méthode DT est préférable à la méthode Monte-Carlo. Plus précisément, comme ces deux méthodes se distinguent uniquement par la façon dont elles évaluent les paires état-action*, je considère que, dans le cadre d'épisodes relativement longs, la méthode DT d'évaluation est préférable à la méthode Monte-Carlo d'évaluation.

Pour évaluer un état ou une paire état-action*, la méthode Monte-Carlo doit attendre le retour. Cela signifie que l'agent doit attendre la fin de l'épisode. Or, certains épisodes peuvent être très longs. Donc, la mise à jour de l'état (ou de la paire état-action*) peut être très lente. De plus, si l'épisode est très long et qu'il faut attendre sa conclusion pour mettre à jour les états (ou les paires état-action*), il faut que l'agent se rappelle de tous les états (ou paires état-action*) qui ont été rencontrés au cours de l'épisode ainsi que de leur retour respectif. Or, peut-être que, en intelligence artificielle, cette capacité de mémorisation ne pose

pas de problème, mais cela pose problème du point de vue de la psychologie, et c'est ce dernier point de vue qui m'intéresse. Enfin, il semble clair que, du point de vue de la psychologie, nous évaluons nos états (ou paires état-action*), non pas en attendant le terme de l'épisode (la fin du retour), mais en se fiant à la valeur de l'état suivant. Par exemple, en quittant le bureau en auto, si j'estime que le retour à la maison prendra 10 minutes, et que, cinq minutes plus tard, étant pris dans le trafic, j'estime qu'il me faudra encore 15 minutes pour me rendre à la maison, alors il est clair que je n'attendrai pas mon arrivée à la maison pour mettre à jour l'estimé que j'ai pris au début du voyage; je vais mettre à jour ce premier estimé sur la base du second estimé[80].

Mais il est clair que, si les épisodes sont très courts, la méthode Monte-Carlo est préférable à la méthode DT. D'une part, comme les épisodes sont très courts, notre capacité à mémoriser le retour des paires état-action* ne pose pas de problème. D'autre part, lorsque l'épisode est très court, nous sommes plus enclins à évaluer une paire état-action* à partir de son retour plutôt qu'à partir de ses récompenses immédiates et de la valeur estimée de la paire état-action suivante. En fait, la question du choix entre méthode DT et méthode Monte-Carlo ne se pose même pas lorsque les paires état-action* mènent *directement* à un état final. Comme l'état final donne immédiatement le retour, il est clair qu'une action qui se déroule dans des épisodes de ce type (des épisodes qui se terminent en une étape) va évaluer les paires état-action* à partir de leur(s) retour(s). Donc, dans ce dernier cas de figure, c'est la méthode Monte-Carlo qui sera utilisée.

8.5 Retour sur notre propre problème

Donc, selon moi, lorsqu'il est question d'épisodes relativement longs, c'est la méthode DT qui rend compte de l'apprentissage permettant à l'agent de percevoir les bonnes affordances. Essayons maintenant de préciser la forme qui doit être donnée à cette méthode d'apprentissage lorsque l'agent est un système perceptuo-attentionnel dont le but est d'apprendre à percevoir les bonnes affordances.

---

[80] Dans cet exemple, le nombre de minutes peut correspondre à la valeur de l'état. Cet exemple est inspiré d'un exemple de Sutton et Barto (1998).

Comme on l'a dit, la première chose que l'agent (le système perceptuel ou, plus précisément, le système perceptuo-attentionnel) doit faire, c'est évaluer sa politique. Comme je l'ai montré, il évalue sa politique en évaluant les paires état-action*. Ici, une paire état-action* est une décision de porter attention à une affordance lorsque le système perceptuel est confronté à un ensemble donné de propriétés perceptibles, au sein desquelles on retrouve des affordances. D'abord, ce que l'algorithme présenté plus haut nous dit, c'est qu'il faut initialiser arbitrairement la valeur de chacune des paires état-action*. Ensuite, pour trouver la valeur réelle des paires état-action*, l'agent doit interagir avec son environnement. Le système perceptuel doit donc faire l'expérience d'une première paire état-action*; autrement dit, sur la base de sa politique, il prend la décision de porter attention à une affordance (disons l'affordance « amortis-moi » présentée par la balle de tennis) à partir de l'état dans lequel il se trouve, qui comprend un ensemble d'affordances. Cette décision de porter attention à l'affordance « amortis-moi » à partir d'un ensemble donné d'affordances le conduit dans un autre état (un autre ensemble d'affordances), où l'agent prend une autre décision. Par exemple, la nouvelle situation pourrait se présenter avec l'affordance « retourne en fond de terrain » et l'affordance « reste au filet », et l'agent pourrait poser l'action* ou prendre la décision de porter attention à l'affordance « retourne en fond de terrain ». C'est à partir de l'expérience de cette deuxième paire état-action* que l'agent peut mettre à jour la valeur de la première paire état-action*. L'assignation présentée plus haut nous prescrit ce qu'il faut faire pour mettre à jour cette première paire état-action*. D'abord, il faut se rappeler que, conformément à l'initialisation des valeurs, la deuxième paire état-action* (c'est-à-dire, l'action* ou la décision de porter attention à l'affordance « retourne en fond de terrain » lorsque l'agent est confronté aux affordances « retourne en fond de terrain » et « reste au filet ») possède une valeur. L'agent additionne alors cette valeur initiale à la récompense (possiblement nulle) qui a été donnée à l'agent lorsqu'il est tombé dans le deuxième état. Formellement, l'assignation présentée plus haut exprime cette somme en ces termes : $r + \gamma Q(s', a')$. Ensuite, conformément à l'assignation « $Q(s, a) \leftarrow Q(s\ a) + \alpha(r + \gamma Q(s', a') - Q(s, a))$ », l'agent soustrait la valeur initiale de la première paire état-action* à cette somme et il additionne cette différence à la valeur initiale de la première paire état-action* (la décision de porter attention à l'affordance « amortis-moi » lorsque l'agent est confronté au premier ensemble d'affordances). L'agent procède de la même façon pour la

deuxième paire état-action*, c'est-à-dire pour la décision de porter attention à l'affordance « retourne en fond de terrain » lorsqu'il est confronté au deuxième ensemble d'affordances, c'est-à-dire aux affordances « retourne en fond de terrain » et « reste près du filet ». L'agent continue ainsi jusqu'au terme de l'épisode, qui correspond à la satisfaction de l'intention ou à un état de l'environnement qui rend impossible la satisfaction de cette intention. L'agent recommence ensuite un autre épisode en procédant de la même façon pour chaque paire état-action*, c'est-à-dire pour chaque décision de porter attention à une affordance donnée à partir d'un ensemble donné d'affordances. L'expérience d'un très grand nombre d'épisodes devrait permettre à l'agent de trouver la valeur réelle des paires état-action*.

La deuxième étape est d'améliorer la politique, c'est-à-dire améliorer un « mapping » d'ensembles d'affordances sur des décisions de faire attention à une affordance donnée. Si l'on suppose que, initialement, la politique de l'agent est stochastique, il existe plusieurs paires état-action* faisant référence à un même état. Autrement dit, face à un même ensemble d'affordances (un même état), plusieurs actions* (c.-à-d., des décisions de faire attention à une affordance) peuvent être prises. Chacune de ces paires état-action* a été évaluée au cours de la première étape, c'est-à-dire au cours de l'évaluation de la politique. Au cours de la deuxième étape, l'amélioration de la politique, l'agent choisit la meilleure paire état-action*. Autrement dit, il choisit la paire « ensemble donné d'affordances-décision de faire attention à telle affordance plutôt qu'aux autres » qui possède la plus grande valeur. Aussi, dans la mesure où il est nécessaire d'avoir une politique stochastique, l'agent peut choisir d'autres paires état-action*. La nouvelle politique (améliorée) de l'agent donnera un degré de probabilité moindre aux paires état-action* qui possèdent moins de valeur.

Après avoir amélioré sa politique, l'agent doit évaluer sa politique améliorée de la façon qui a été décrite ci-dessus. Ensuite, l'agent doit de nouveau l'améliorer. Une nouvelle politique, encore meilleure, est alors formée. Et les deux mêmes étapes sont répétées pour cette nouvelle politique ainsi que pour toutes les autres politiques qui résultent de l'application de ces deux étapes. Grâce à ce processus itératif, il y a convergence vers une politique optimale, c'est-à-dire, pour nous, convergence vers une politique qui fait en sorte que, lorsque l'agent est confronté à un ensemble donné d'affordances, cet agent fait attention à l'affordance qui lui permet de satisfaire au mieux son intention.

Jusqu'ici, nous nous sommes concentrés sur l'apprentissage qui est à la base de notre capacité à faire attention à la meilleure affordance. Or, comme je l'ai dit au début du chapitre, l'apprentissage par renforcement n'est pas seulement à la base de cette capacité; il est aussi à la base de notre capacité à faire attention à la meilleure variable optique de contrôle. C'est de ce dernier point dont j'aimerais m'assurer avant de passer au chapitre suivant.

Pour que notre capacité à faire attention à la bonne variable optique de contrôle puisse être conçue comme le fruit d'un apprentissage par renforcement, il suffit que l'on puisse déterminer les cinq éléments suivants : l'agent, les actions* de l'agent, les états de l'environnement, les politiques et les récompenses. Il est clair qu'il est possible de déterminer ces éléments.

Encore une fois, l'agent est le système perceptuel. Les actions* de l'agent correspondent aux décisions de faire attention à une variable optique. Les états de l'environnement correspondent aux ensembles de variables optiques. Les politiques correspondent aux « mappings » qui associent un ensemble de variables optiques à une décision de faire attention à telle variable optique plutôt qu'aux autres. Et des récompenses peuvent être données à l'agent à chaque fois que l'action* cause un état de l'environnement qui correspond à ce qui est prescrit par l'affordance.

La description qui a été donnée ci-dessus et qui nous montre comment l'agent apprend à percevoir les bonnes affordances fait référence à une politique qui s'insère au sein d'épisodes relativement longs. Mais, à l'exception d'un cas important, sur lequel nous reviendrons plus tard, les problèmes d'apprentissage par renforcement qui nous intéressent ne font pas intervenir de longs épisodes; nous pouvons même dire que les politiques de nos agents font intervenir des actions* qui mène directement à l'état final. Les considérations suivantes devraient nous en convaincre.

Les perceptions d'affordances (les actions*) que l'agent (le système perceptuel) doit retenir dans sa politique optimale sont sous le contrôle d'intentions de très bas niveaux. L'action* (la décision de faire attention à telle affordance plutôt qu'aux autres) de l'agent doit mener à l'état environnemental auquel réfère cette intention de bas niveau. Autrement dit, cet état environnemental est l'état final de l'épisode. Or, une intention de bas niveau réfère à l'état environnemental qui suit immédiatement l'état actuel. Donc, pour atteindre l'état final

de l'épisode, l'agent n'a pas besoin de poser plus d'une action* (une décision de faire attention à telle affordance plutôt qu'aux autres). Les épisodes sont donc minimalement courts. L'apprentissage par renforcement par lequel l'agent apprend à percevoir la bonne affordance est donc de type Monte-Carlo.

Le même constat vaut pour l'apprentissage par renforcement qui est à la base de notre capacité à percevoir la bonne variable optique : l'état environnemental qui fait suite à l'action* (la décision de faire attention à cette variable optique plutôt qu'aux autres) est toujours l'état final de l'épisode. En effet, le système perceptuel responsable de la perception de la variable optique (ou action*) est sous le contrôle de la perception de l'affordance et le mouvement auquel réfère la perception de l'affordance est l'état final auquel doit mener l'action* (la décision de porter attention à cette variable optique plutôt qu'aux autres); or, le mouvement auquel réfère la perception de l'affordance est toujours atteint par *une seule* action*. Donc, encore une fois ici, c'est par la méthode Monte-Carlo que le système perceptuel va apprendre à percevoir la bonne variable optique.

Ce qui précède pourrait laisser l'impression que la méthode DT, qui est préférable lorsque la politique est insérée au sein d'épisodes relativement longs, a été décrite pour rien. Comme nous allons le voir, la méthode d'apprentissage DT va nous servir pour expliquer le changement qui caractérise le passage à la réactivité.

Pour finir, j'aimerais souligner ceci. Jusqu'ici, nous avons vu comment l'organisme apprend à adopter la bonne loi de contrôle, comment il apprend à percevoir la bonne variable optique de contrôle et comment il apprend à percevoir la bonne affordance. Pour avoir une vision plus globale et plus intuitive des choses, il faudrait voir comment ces trois apprentissages s'articulent entre eux. C'est dans la conclusion de la présente thèse que je décrirai l'interaction qui existe entre ces différents apprentissages.

8.6 Résumé

Les méthodes d'apprentissage par renforcement permettent de résoudre un problème : Comment, à travers ses interactions avec son environnement, l'agent (pour nous : le système perceptuo-attentionnel) peut-il en arriver à adopter une politique (pour nous : un ensemble d'associations entre des ensembles d'affordances et des décisions de faire attention à une

affordance) optimale, c'est-à-dire une politique qui procure un maximum de récompenses. Trois méthodes permettant de résoudre ce problème ont été présentées : la méthode PD, la méthode Monte-Carlo et la méthode DT.

Ces trois méthodes se rejoignent en ce qu'elles procèdent d'abord à une évaluation de la politique actuellement suivie. Sur la base de cette évaluation, elles améliorent ensuite la politique actuellement suivie. La nouvelle politique fait ensuite l'objet d'une nouvelle évaluation et, sur la base de cette nouvelle évaluation, d'une nouvelle amélioration. Il faut répéter ces deux étapes (évaluation et amélioration) sur chaque politique améliorée. En procédant de la sorte, l'agent devrait pouvoir établir une politique optimale. La politique sera optimale lorsque la politique « améliorée » sera identique à la politique précédente.

Ces trois méthodes divergent quant à la première étape, l'étape de l'évaluation. Dans ce qui suit, je vais résumer la méthode d'évaluation de chacune de ces trois méthodes.

La méthode d'évaluation PD des états de l'environnement procède à des « balayages » itératifs où, lors d'une itération, chaque état est évalué sur la base de tous ses états successeurs possibles et sur leur probabilité. Plus précisément, l'évaluation d'un état, lors d'une itération, observe l'équation suivante : $V_{k+1}(s) = \sum_a \pi(s,a) \sum_{s'} P_{s\,s'}^{a} [\Re_{s\,s'}^{a} + \gamma V_k(s')]$.

Comme je l'ai dit, l'évaluation d'une politique est suivie par son amélioration. Pour améliorer une politique, la méthode PD observe l'état (ou les états) qui succède(nt) à une action* posée à partir d'un état s. La méthode PD conserve le(s) action(s)* qui, à partir d'un état s donné, donne(nt) le plus de récompenses immédiates et qui est (sont) suivie(s) par le(s) état(s) ayant le plus de valeur. Ensuite, comme je l'ai dit plus haut, il faut de nouveau évaluer la nouvelle politique améliorée et améliorer cette dernière. En répétant ces deux étapes sur chaque nouvelle politique améliorée, on devrait converger vers la politique optimale.

La méthode d'évaluation Monte-Carlo procède en faisant interagir l'agent avec son environnement. Grosso modo, l'évaluation de l'état s consiste à faire la moyenne de tous les retours qui suivent l'état s. Et, pour l'évaluation d'une paire état-action* (s,a), il faut faire la moyenne des retours qui suivent l'état s, à partir duquel l'action* a a été exécutée. L'évaluation des états et des paires état-action* peut aussi s'appuyer sur une méthode

incrémentielle : $V(s_t) \leftarrow V(s_t) + \alpha(R_t - V(s_t))$, pour l'évaluation des états, et $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(R_t - Q(s_t, a_t))$, pour l'évaluation des paires état-action*.

Pour améliorer la politique actuellement suivie, la méthode Monte-Carlo s'appuie sur l'évaluation des paires état-action*. Cela tient au fait que la méthode Monte-Carlo ne dispose pas d'un modèle complet du monde; la méthode Monte-Carlo ne permet pas de savoir quel état s' se produira après qu'il aura exécuté l'action* a à partir de l'état s. En effet, la méthode Monte-Carlo ne dispose pas des probabilités $P^{a}_{s\,s'}$. De plus, soulignons que, pour améliorer sa politique, la politique ne doit pas être déterministe. Et si elle est déterministe, on doit poser l'existence de débuts exploratoires.

La méthode d'évaluation DT procède également sur la base, non pas d'un modèle complet du monde, mais sur la base d'expériences avec l'environnement. Aussi, comme elle ne dispose pas d'un modèle complet du monde, la méthode TD ne peut pas améliorer la politique actuellement suivie en se fiant à une évaluation des états; elle doit se fier à l'évaluation des paires état-action*. Cependant, contrairement à la méthode Monte-Carlo, elle n'évalue pas les paires état-action* en calculant leurs retours. Pour évaluer la paire état-action* (s,a), la méthode TD tient compte de la récompense immédiate qui suit l'action* a exécutée à partir de l'état s et elle tient compte de la valeur de la prochaine paire état-action* (s',a'). Plus précisément, on procède de manière incrémentielle pour calculer la valeur de la paire état action* (s,a). C'est appliquant itérativement l'assignation suivante que l'on calcule la valeur de la paire état-action* (s,a) : $Q(s, a) \leftarrow Q(s\ a) + \alpha(r + \gamma Q(s', a') - Q(s, a))$.

Ensuite, la méthode DT parvient à trouver la politique optimale en procédant comme pour la méthode PD et la méthode Monte-Carlo.

CHAPITRE IX

APPRENDRE À RÉAGIR

Selon notre modèle, lorsque l'agent devient expert, les intentions de bas niveau sont éliminées et ce sont les intentions du niveau immédiatement supérieur qui prennent en charge le contrôle des perceptions d'affordances, qui étaient jusqu'alors prises en charge par les intentions du niveau inférieur. Ainsi, si le plan possède trois niveaux, le troisième niveau, le niveau le plus bas, est éliminé au profit des intentions du second niveau; ce sont maintenant ces intentions qui contrôlent le système perceptuel. Et, si un degré d'expertise additionnel est possible, les intentions du second niveau seront éliminées au profit de l'intention du premier niveau; c'est maintenant cette intention qui va contrôler le système perceptuel. De la sorte, le système devient de plus en plus réactif : l'action de l'agent est de moins en moins contrôlée par un plan et la réactivité aux affordances prend presque toute la place[81]. Maintenant, existe-il un mécanisme d'apprentissage qui puisse expliquer ces changements au sein de la structure de contrôle du système perceptuel? Je soutiens que oui, et j'espère pouvoir le montrer dans ce qui suit.

J'ai divisé le chapitre 9 en deux grandes parties. Dans la première partie, je présente ce mécanisme d'apprentissage; dans la seconde partie, je réponds à des objections.

---

[81] La description qui vient d'être donnée simplifie énormément les choses. Dans cette description, les intentions correspondant à un niveau du plan sont toutes éliminées simultanément. Il est plus plausible de penser que les intentions correspondant à un niveau du plan sont éliminées à des moments différents. Par exemple, si le joueur de tennis a plus d'aptitudes pour le jeu offensif, les intentions particulières subordonnées à l'intention « jouer offensivement » devraient vraisemblablement être éliminées avant les intentions particulières qui sont subordonnées à l'intention « jouer défensivement ». Cependant, pour simplifier ma description, je vais supposer dans ce qui suit que les intentions correspondant à un niveau du plan sont toutes éliminées simultanément.

9.1 Le mécanisme d'apprentissage

Tout d'abord, il faut préciser la nature des changements qui se produisent lorsque le système perceptuel passe d'un état où il est contrôlé par un ensemble très chargé d'intentions à un état où il est contrôlé par un petit groupe d'intentions (voire une seule intention). Mon point est le suivant : d'un point de vue computationnel, les changements les plus intéressants se produisent, non pas au niveau de la structure des intentions, mais au niveau du système perceptuel. Supposons que, initialement, le système perceptuel soit contrôlé par trois intentions de bas niveau. Et supposons que, à l'issue d'un processus d'apprentissage qu'il reste à déterminer, c'est maintenant une intention de niveau supérieur qui contrôle le système perceptuel de l'agent. Maintenant, une seule intention contrôle le système perceptuel là où il y en avait trois initialement. Dans les deux situations, le système perceptuel doit émettre les *mêmes* outputs. Plus précisément, le système perceptuel doit émettre les mêmes perceptions d'affordances, qu'il soit confronté à trois inputs (trois intentions) ou à un input (une intention). De plus, il ne faudrait pas croire que cet input unique contient la même information que l'information contenue au sein des trois inputs. Cet input unique correspond à une intention et les trois inputs correspondent à des intentions qui sont des moyens relativement à la première intention. Or, l'information qui m'est donnée lorsque je connais une fin et l'information qui m'est donnée lorsque je connais les moyens pour atteindre cette fin ne sont pas la même information. Donc, étant donné que le système perceptuel émet les mêmes outputs lorsqu'il est soumis à une information qui est différente d'une fois à l'autre, le système perceptuel (ses capacités computationnelles) *change* considérablement lors de l'apprentissage (lors de la diminution du plan). Or, le chapitre 8 a montré que les changements d'ordre perceptuel étaient le fruit d'un apprentissage par renforcement. Donc, il est probable que les changements qui se produisent lorsque l'agent devient réactif soient le fruit d'un apprentissage par renforcement, comme celui qui a été présenté lors du chapitre 8. C'est précisément ce que je vais montrer.

Mais avant de montrer comment les choses se passent, il faut rappeler que, au moment d'éliminer une ou plusieurs intentions au profit d'une intention supérieure, l'agent est déjà un planificateur exemplaire. Autrement dit, l'agent produit des plans maximalement efficaces. Cela signifie que les intentions que l'agent génère sont les meilleures qui soient. Et j'ajoute

que, au moment d'éliminer des intentions au profit d'une intention supérieure, l'agent est en mesure de percevoir les affordances qui permettent de réaliser les intentions du bas de la hiérarchie. Autrement dit, l'agent a su, pour chaque intention du bas de la hiérarchie, établir une politique perceptuelle qui satisfait l'intention. C'est grâce à l'apprentissage perceptuel décrit au chapitre 8 qu'il a pu établir ces politiques perceptuelles. Mais, ici, il faut se donner une idée plus claire de ces intentions du bas de la hiérarchie et de la politique perceptuelle qui est contrôlée par chacune de ces intentions.

Les intentions du bas de la hiérarchie sont certainement plus spécifiques que les intentions situées plus haut dans la hiérarchie. Cependant, les intentions du bas de la hiérarchie ne sont pas pour autant entièrement déterminées; elles possèdent encore un caractère typique ou général. Autrement dit, ces intentions possèdent un contenu conceptuel[82]. Leur contenu conceptuel fait référence à un mouvement typique (non déterminé) et à un changement environnemental typique (non déterminé). Ce changement environnemental typique correspond au changement environnemental typique qui se produit à la suite de l'exécution du mouvement typique. Ainsi, si un joueur de tennis a l'intention de jouer un coup droit dans le coin supérieur droit du terrain, cette intention du bas de la hiérarchie représenterait, d'une part, un mouvement typique (un coup droit) et, d'autre part, le changement que subit l'environnement typique à la suite de ce mouvement typique. Par exemple, l'agent se représenterait (de manière conceptuelle) la balle dans le coin droit supérieur droit du terrain. Mais il y a plus : l'état *initial* de l'environnement typique est aussi représenté dans l'intention du bas de la hiérarchie. À quoi correspond cet environnement typique initial? Comme on l'a dit au chapitre 1, l'intention du bas de la hiérarchie est produite lorsque l'agent fait face à un environnement donné. C'est cet environnement initial, considéré dans sa typicité, qui est représenté dans l'intention. Cependant, cet environnement typique initial est considéré dans l'intention comme état environnemental initial, pas comme état environnemental final. Je suis conscient du fait que l'attribution d'un environnement typique initial à l'intention n'est pas

[82] À ce titre, les intentions du bas de la hiérarchie se distinguent des intentions en action de Searle (1983). En effet, chez Searle, les intentions en action possèdent un contenu *non* conceptuel. De plus, contrairement aux intentions en action de Searle, les intentions du bas de la hiérarchie *représentent* le mouvement; elles ne le *présentent* pas. Cela découle du fait que, selon mon modèle, l'intention du bas de la hiérarchie n'est pas une condition suffisante pour l'exécution du mouvement. Selon mon modèle, l'exécution du mouvement requiert, en plus de l'intention du bas de la hiérarchie, une perception d'affordance et une perception de variable optique.

habituelle dans la littérature. Cependant, comme cette attribution n'est pas trop contre-intuitive et comme elle me rendra un grand service pour la suite des choses, je me suis autorisé cette caractérisation du contenu de l'intention.

Maintenant, le lecteur pourrait s'interroger sur la pertinence d'attribuer un caractère typique au contenu de l'intention. Il faut voir que cette attribution d'un contenu typique aux intentions du bas de la hiérarchie est requise pour qu'il puisse y avoir apprentissage. Plus précisément, si les intentions du bas de la hiérarchie avaient un contenu entièrement déterminé (non typique), l'apprentissage des politiques perceptuelles associées à ces intentions serait beaucoup trop long. En effet, si ces intentions avaient un contenu parfaitement déterminé, il y aurait une quasi-infinité d'intentions du bas de la hiérarchie. Or, s'il y avait une quasi-infinité d'intentions, l'apprentissage de la totalité des politiques associées à ces intentions serait (quasiment) infiniment long. Ce n'est pas le cas si les intentions se voient attribuer un contenu typique. En effet, dans le cas du tennis par exemple, il existe probablement quelques dizaines d'intentions typiques optimales[83]. Donc, le nombre d'intentions optimales est plutôt limité, ce qui rend possible l'apprentissage des politiques que ces intentions contrôlent. Penchons-nous maintenant sur la politique contrôlée par une intention du bas de la hiérarchie.

Une politique, comme on l'a dit au chapitre précédent, est un ensemble d'associations entre des états environnementaux et des actions* (c.-à-d. : des décisions de faire attention à telle affordance plutôt qu'aux autres.). Quels sont donc les états environnementaux qu'une politique associée à une intention du bas de la hiérarchie représente? L'intention située au bas de la hiérarchie spécifie déjà cet état environnemental. En effet, comme on vient de le dire, l'intention du bas de la hiérarchie répond à une situation environnementale et cette situation environnementale est représentée dans l'intention en tant que situation environnementale initiale. C'est cette situation environnementale qui est représentée dans la politique contrôlée par l'intention du bas de la hiérarchie. Mais, comme on l'a dit, c'est une situation environnementale typique que l'intention représente. La question se pose donc de savoir si la politique fait référence à cet environnement typique ou si, au contraire, elle fait référence à tous les états environnementaux particuliers ou déterminés qui peuvent réaliser ou instancier

[83] Je rappelle que, à ce stade, l'agent est un planificateur exemplaire. Par conséquent, l'agent ne possède que des intentions optimales.

cet état environnemental typique. Il faut que la politique perceptuelle associée à l'intention du bas de la hiérarchie fasse uniquement référence à l'état environnemental typique, celui auquel fait référence l'intention. Cela découle du fait que, si la politique perceptuelle faisait référence à tous les états environnementaux qui instancient l'état environnemental typique, l'apprentissage serait beaucoup trop long. En effet, l'état environnemental typique peut être réalisé à travers une quasi-infinité d'états environnementaux particuliers ou déterminés. Donc, l'apprentissage de la politique perceptuelle optimale prendrait un temps beaucoup trop considérable. Nous en savons donc davantage sur la politique perceptuelle associée à une intention du bas de la hiérarchie. Cette politique représente un « mapping » entre un état environnemental typique et quelques actions* (des décisions de faire attention à une affordance plutôt qu'aux autres). En effet, initialement, c'est-à-dire avant le terme de l'apprentissage, cette politique est stochastique. Elle associe à l'état environnemental typique quelques actions* et chacune de ces actions* reçoit une probabilité donnée. C'est l'apprentissage perceptuel décrit au chapitre précédent qui fait en sorte que, à l'issue de cet apprentissage, une seule de ces actions* est associée à l'état environnemental typique. Autrement dit, l'apprentissage perceptuel fait en sorte que cette action* reçoive une probabilité de 1. Pourquoi cet apprentissage conduit-il, non pas à l'instauration de plusieurs actions*, mais à l'instauration d'une seule action*? Parce que l'intention du bas de la hiérarchie est très spécifique, de sorte qu'une seule affordance est en mesure de la satisfaire.

Il faut encore souligner ceci : l'apprentissage d'une politique contrôlée par une intention du bas de la hiérarchie ne requiert pas une grande période d'apprentissage. L'argument à la base de cette dernière proposition est le suivant : Si une politique produit des épisodes qui n'ont pas plus d'une étape, la période d'apprentissage requise pour donner un caractère optimal à cette politique est courte. Or, les politiques qui sont contrôlées par des intentions du bas de la hiérarchie ont des épisodes qui n'ont pas plus d'une étape. À la lumière du chapitre 8, où il a été question de l'apprentissage par renforcement, nous savons que la première prémisse de l'argument ci-dessus est évidente. Mais qu'est-ce qui nous permet de dire que les politiques contrôlées par des intentions du bas de la hiérarchie produisent des épisodes d'actions qui n'ont pas plus d'une étape? Un épisode d'actions prend fin lorsqu'une action* conduit à un état environnemental qui ou bien satisfait l'intention du bas de la hiérarchie, ou bien rend impossible la satisfaction de cette intention. Or, les intentions du bas de la

hiérarchie possèdent un caractère immédiat et ce caractère immédiat fait en sorte qu'il suffit d'une *seule* action* pour atteindre un état environnemental final, c'est-à-dire un état environnemental qui ou bien satisfait l'intention, ou bien rend impossible la satisfaction de l'intention.

J'en ai maintenant assez dit sur les intentions du bas de la hiérarchie et sur leur politique perceptuelle. Il est maintenant temps de voir ce qui se passe lorsque ces intentions sont éliminées au profit d'une intention supérieure. Supposons que le joueur de tennis est habité par l'intention *supérieure* de jouer de manière offensive. Comme nous supposons que ce joueur est déjà un planificateur exemplaire, celui-ci réalise cette intention supérieure à travers des intentions inférieures optimales. Il s'agit maintenant d'éliminer toutes ces intentions inférieures optimales et de ne garder que l'intention supérieure, l'intention de jouer offensivement. L'élimination des intentions inférieures ne sera pas dommageable pour les performances de l'agent à condition que les perceptions d'affordances qui étaient causées par les intentions inférieures soient aussi causées par l'intention supérieure. Or, les intentions inférieures causaient ces perceptions d'affordances en contrôlant leur politique perceptuelle. Donc, la question importante est celle-ci : comment éliminer les intentions inférieures au profit d'une intention supérieure, cela sans que les politiques perceptuelles des intentions inférieures soient perdues? À cette question, je réponds : en « fusionnant » les politiques inférieures en une politique englobante. Plus précisément, j'énonce l'hypothèse suivante : lorsque des intentions inférieures sont éliminées, leurs politiques sont « fusionnées » en une politique englobante et cette politique englobante est dorénavant contrôlée par l'intention supérieure. Pour que la performance de l'agent soit conservée, il faut que, lorsqu'un environnement donné survient, la politique englobante émette l'action* (la perception d'affordance) que la politique inférieure émettait lorsqu'elle était confrontée à ce même environnement. La politique englobante permet précisément cela. Mais définissons d'abord la fusion.

Lorsque des politiques sont « fusionnées » et que, donc, une politique englobante est formée, tous les états environnementaux présents au sein des politiques inférieures sont rassemblés au sein d'une même politique (la politique englobante) et on associe à chacun de ces états environnementaux typiques les actions* que les politiques inférieures leur

associaient. Donc, dans notre exemple du joueur de tennis qui est en voie d'éliminer les intentions inférieures au profit de l'intention de jouer offensivement, cela signifie que toutes les situations typiques qui se prêtent à un coup offensif et qui étaient distribuées au sein des politiques inférieures sont rassemblées au sein d'une même politique et il faut associer à ces situations les actions* que les politiques inférieures leur associaient.

Mais si la « fusion » de politiques inférieures consiste simplement à rassembler les situations typiques au sein d'une politique plus englobante et à associer à ces situations typiques les actions* que les politiques initiales leur associaient, il y a tout lieu de se demander si, oui ou non, il est juste de dire que la « fusion » de politiques inférieures en une politique englobante fait appel à un processus d'apprentissage. Or, dans un bon nombre de cas, il y a effectivement apprentissage. Cet apprentissage se produit lorsque, à l'issue de la « fusion », plusieurs actions* sont associées à une *même* situation typique. C'est ce qui se produit lorsque des intentions inférieures font référence à la même situation environnementale *initiale*. Lorsque cela arrive, les politiques inférieures font référence à la même situation environnementale typique. Ce cas de figure se présente, par exemple, si, face à une même situation environnementale typique, l'agent a parfois l'intention (optimale) de jouer un amorti et parfois l'intention (optimale) de jouer un coup en fond de terrain. Dans ce cas de figure, les politiques de ces deux intentions associent une action* différente à une *même* situation typique. Donc, la politique englobante va associer deux actions* différentes à la même situation typique. Pourquoi est-ce problématique? Parce que la politique englobante va associer des probabilités arbitraires à ces deux actions* et parce que, étant donné que ces probabilités seront arbitraires, il y a tout lieu de penser qu'elles ne seront pas optimales. Par exemple, il se pourrait que la politique englobante attribue une probabilité de 0.5 à l'action* de faire attention à l'affordance « il est possible de jouer un coup en fond de terrain » et une probabilité de 0.5 à l'action* de faire attention à l'affordance « il est possible de jouer un amorti ». Tout joueur de tennis sait que ces probabilités ne sont pas optimales. Au contraire, il serait plus plausible de penser que, pour avoir un comportement optimal, il faille attribuer une probabilité de 0.9 à l'affordance « il est possible de jouer un coup en fond de terrain » et une probabilité de 0.1 à l'affordance « il est possible de jouer un amorti ». L'apprentissage doit donc nous amener à attribuer des probabilités optimales aux actions*. L'apprentissage qui nous permet d'attribuer des probabilités optimales aux actions* est un apprentissage par

renforcement, le même type que celui dont il a été question dans le chapitre précédent. La façon dont l'apprentissage par renforcement permet d'attribuer des probabilités optimales aux actions* n'a cependant pas été décrite au chapitre précédent. Je me permets donc dans ce qui suit de décrire comment l'apprentissage par renforcement permet d'attribuer des probabilités optimales aux actions*.

Initialement, c'est-à-dire lorsque les probabilités attribuées aux actions* sont arbitraires et sous-optimales, chaque action* peut conduire à des états environnementaux différents et ces états environnementaux donnent des récompenses différentes. Par exemple, lorsqu'elle reçoit une probabilité de 0.5, l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un amorti » mène quelquefois à un état environnemental qui procure un maximum de récompenses (l'amorti fait gagner le point à l'agent) et quelquefois à un état environnemental qui procure un minimum de récompense (l'amorti permet à l'adversaire de jouer un coup gagnant). Lorsque la probabilité attribuée à l'action* de faire attention à l'affordance « il est possible de jouer un amorti » est arbitraire (disons 0.5), cette action* va souvent mener à l'état environnemental qui procure un minimum de récompense. Cela signifie que cette paire « état-action* » se verra attribuer une petite valeur[84]. Cela est un signe que la probabilité arbitrairement attribuée à cette action* est sous-optimale. C'est ainsi qu'il faudra changer la probabilité de cette action*. Plus précisément, une nouvelle distribution de probabilités sera essayée. Par exemple, l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un coup en fond de terrain » se verra attribuer la probabilité 0.6 et l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un amorti » se verra attribuer une probabilité de 0.4. Après avoir changé cette distribution des probabilités, les actions* conduiront à des états environnementaux qui n'auront plus la même fréquence. Ainsi, comme cette nouvelle distribution des probabilités fera en sorte que mon adversaire sera davantage surpris par mon amorti, l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un amorti » conduira plus souvent qu'avant à l'état environnemental

[84] Cela est conditionnel à ce que l'état environnemental auquel mène l'action* soit un état environnemental final. Si cet état environnemental n'était pas final, le calcul de la valeur de la paire « état-action » serait plus compliqué. Il faudrait, en plus de la récompense que procure l'état s', tenir compte de la *valeur* de la paire « état-action » qui se réfère à l'état s' et à l'action a'. Cependant, pour simplifier ma description, je vais supposer ci-dessus que l'état environnemental auquel mène l'action* est un état environnemental final.

qui donne un maximum de récompenses (un amorti gagnant). Donc, cette paire « état-action* » gagnera en valeur par cette nouvelle distribution des probabilités. Mais, comme tout joueur de tennis le sait, cette distribution des probabilités n'est toujours pas optimale. Il faudra de nouveau changer la distribution des probabilités et voir si les actions* auxquelles on aura attribué de nouvelles probabilités conduiront plus souvent à des états environnementaux qui procurent un maximum de récompenses. Et c'est en continuant à procéder de la sorte que l'agent devrait atteindre une distribution de probabilités optimales. Lorsqu'une telle distribution des probabilités est atteinte, il n'est plus possible de trouver une distribution de probabilités qui donne aux paires « état-action* » davantage de valeur. Cela signifie entre autres que l'action* qui est effectuée à partir de l'état s et qui consiste à faire attention à l'affordance « il est possible de faire un amorti » mènera la plupart du temps à un état environnemental qui procure un maximum de récompenses (par ex., un coup gagnant). Dans ce cas de figure, on peut par exemple penser que l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un coup en fond de terrain » se verra attribuer la probabilité 0.9 et que l'action* qui consiste à faire attention à l'affordance « il est possible de jouer un amorti » se verra attribuer la probabilité 0.1.

À la suite de cet apprentissage et, donc, à la suite de la « fusion » de politiques perceptuelles en une politique englobante optimale, l'action de l'agent a changé de nature. L'agent ne base plus son action sur un plan détaillé; il réagit beaucoup plus directement aux affordances. Il n'a pas besoin de générer des intentions de bas niveau avant de réagir aux affordances; l'intention de jouer offensivement lui suffit pour réagir aux affordances. Cela signifie que, devant une situation donnée, qui contient une multitude d'affordances, l'agent ne se trouve pas confronté à un choix d'affordance où il aurait à peser le pour et le contre de chaque affordance et où, sur la base de cette délibération, il aurait à produire une intention inférieure (une intention située au bas de la hiérarchie). Au contraire, l'agent, à partir d'une intention supérieure (par ex., jouer de manière offensive), réagit directement à la bonne affordance sans avoir à réfléchir.

La description qui vient d'être donnée du passage à la réactivité peut susciter des objections. Je me tourne maintenant vers ces objections.

9.2. Quelques objections

Comme on l'a vu, la politique englobante possède un caractère stochastique; dans cette politique, plusieurs actions* peuvent être associées à une même situation environnementale et des probabilités sont attribuées à ces actions*. Donc, le fait que, en face d'une situation environnementale donnée, l'agent perçoive une affordance plutôt qu'une autre et que, donc, il choisisse une action plutôt qu'une autre, repose sur une probabilité. Or, on peut s'opposer à cette conséquence de mon modèle du passage à la réactivité. L'argument à la base de cette objection est le suivant. Si la sélection de l'affordance et de l'action repose sur une probabilité, alors la *pertinence* de cette affordance et de cette action n'est pas assurée. Or, l'affordance perçue et, donc, l'action de l'expert sont généralement pertinentes. Donc, l'expert ne sélectionne pas son action (et son affordance) à partir d'une probabilité.

Pour invalider cet argument, je vais montrer que mon modèle du passage à la réactivité n'implique pas la proposition conditionnelle de l'argument précédent. Plus précisément, je vais montrer que mon modèle du passage à la réactivité implique la proposition conditionnelle suivante : « Si la sélection de l'affordance et de l'action repose sur une probabilité, alors la pertinence de cette affordance et de cette action est généralement assurée. » Et comme mon modèle du passage à la réactivité implique l'antécédent de cette dernière proposition conditionnelle, nous pouvons inférer que mon modèle du passage à la réactivité implique aussi le conséquent de cette proposition. Donc, dans mon modèle, la pertinence de l'affordance et de l'action sélectionnées est généralement assurée. Montrons maintenant que mon modèle du passage à la réactivité permet d'établir la proposition conditionnelle précédente, soit que, si la sélection de l'affordance et de l'action repose sur une probabilité, alors la pertinence de cette affordance et de cette action est généralement assurée.

La sélection de l'affordance et de l'action est pertinente, quand bien même cette sélection repose sur une probabilité, car, dans mon modèle, les différents choix qui s'offrent à l'agent et auxquels des probabilités ont été attribuées sont *tous* pertinents. Dès lors, ce qui me reste à faire pour établir la proposition conditionnelle susmentionnée, c'est de montrer qu'il est vrai que les différents choix qui se présentent à l'agent sont tous pertinents.

Mais montrons de manière plus précise ce que j'entends faire dans ce qui suit. Lors du passage à la réactivité, l'agent a pu se construire une politique englobante, et cette politique possède un caractère stochastique. En effet, comme nous l'avons dit, plusieurs actions*, auxquelles des probabilités ont été attribuées, peuvent être associées à un même état environnemental. Donc, ces actions* sont activées de manière probabiliste. Cela signifie que l'affordance à laquelle fait attention l'agent, ainsi que l'action qui résulte de cette perception, résulte d'un processus probabiliste. Tout le problème est maintenant de montrer que les actions* (et donc les perceptions d'affordances et les actions) de la politique englobante sont *toutes* pertinentes.

Dans la politique englobante, les actions* qui sont associées à un même état environnemental ont été tirées des politiques qui étaient sous le contrôle des intentions du bas de la hiérarchie. Au moment de la fusion, ces intentions inférieures sont optimales. En effet, comme on l'a dit, la fusion et, donc, le passage à la réactivité se produisent uniquement *après* que l'agent soit devenu un planificateur exemplaire. De plus, comme l'agent est déjà passé à travers un processus d'apprentissage, les politiques particulières, contrôlées par ces intentions inférieures, sont optimales. Cela signifie que l'action* de ces politiques permet de réaliser au mieux l'intention inférieure. Et comme, du fait de son optimalité, cette intention inférieure satisfait elle-même au mieux l'intention supérieure, il s'ensuit que l'action* permet aussi de réaliser au mieux l'intention supérieure. Cette action*, ainsi que l'état environnemental auquel elle est associée, sera reprise et intégrée dans la politique englobante, celle qui est contrôlée par l'intention supérieure. Il en sera de même des autres actions*, celles qui se trouvent dans des politiques perceptuelles sous le contrôle d'autres intentions inférieures optimales. Comme ces intentions inférieures sont elles aussi optimales, ces autres actions* permettent aussi de réaliser au mieux l'intention supérieure; elles sont toutes pertinentes relativement à cette intention supérieure. Donc, les actions* qui sont associées au même état environnemental et qui se retrouvent dans la politique englobante permettent *toutes* de réaliser au mieux l'intention supérieure. Donc, lorsque, à partir de l'activation probabiliste d'une action*, l'agent sélectionnera de manière probabiliste une affordance, cette affordance, quelle qu'elle soit, permettra de réaliser au mieux l'intention supérieure. À ce titre, c'est-à-dire en tant qu'elle satisfait au mieux l'intention supérieure, l'affordance sélectionnée (de manière

probabiliste) est toujours *pertinente*. Mais, attention, cette dernière conclusion n'est pas tout à fait juste.

Il arrive que, face à une situation donnée, une action perde de sa pertinence si elle est exécutée trop souvent. Cela découle du fait que si nous ne varions pas suffisamment nos coups, notre adversaire pourra plus facilement les anticiper. Ainsi, si je frappe un amorti avec une probabilité de 0.9, mon adversaire ne sera plus du tout surpris par ce coup. Il faut donc, pour jouer de manière optimale, que j'attribue à l'amorti une probabilité plus petite que celle qui se rapporte au coup en fond de terrain. De même, si mon coup droit est meilleur que mon revers, alors, face à une situation qui sollicite tant un revers qu'un coup droit comme coups pertinents, je devrai jouer plus souvent un coup droit, mais pas uniquement un coup droit, car mon adversaire pourrait alors anticiper mes coups. *Donc, la pertinence d'un coup ne dépend pas seulement de la situation; la pertinence d'un coup est aussi fonction de sa fréquence (sa probabilité d'être exécuté).* Donc, au début de la fusion, c'est-à-dire immédiatement après que les actions* aient été intégrées avec des probabilités *arbitraires* au sein de la politique englobante, les coups ne sont pas encore pertinents au sens prégnant du terme. Mais comme on l'a dit plus haut, l'attribution de probabilités optimales aux actions* peut être atteinte à la suite d'un processus d'apprentissage (un apprentissage par renforcement). Et ce n'est qu'à la suite de ce processus d'apprentissage que l'on peut affirmer la proposition conditionnelle susmentionnée, soit que si la sélection de l'affordance et de l'action repose sur une probabilité, alors cette sélection est pertinente.

Mais est-ce que le processus d'apprentissage qui permet d'attribuer des probabilités optimales aux actions* est un processus d'apprentissage raisonnable? Plus précisément, ce processus d'apprentissage ne prendra-t-il pas un temps trop considérable avant d'arriver à terme? Je ne crois pas. Les intentions du bas de la hiérarchie possèdent un caractère typique. En effet, comme je l'ai dit plus haut, le contenu d'une intention du bas de la hiérarchie fait référence à une situation typique et à un mouvement typique. Les intentions du bas de la hiérarchie sont donc en nombre limité. Cela implique que les actions* qui seront intégrées dans la politique englobante seront aussi en nombre limité. On ne devrait donc pas avoir à associer trop d'actions* à une même situation environnementale typique. Par exemple, on peut penser que, au tennis, il n'y a que deux ou trois coups offensifs optimaux (et, donc, deux

ou trois actions* offensives pertinentes) qui puissent être associés à une même situation environnementale typique. Cela fait en sorte que l'apprentissage permettant d'attribuer des probabilités optimales à ces actions* ne devrait pas prendre un temps si considérable. Mais, de toute façon, il faut voir que, même si cet apprentissage prend un temps plus ou moins considérable, cela ne constitue pas en soi une objection déterminante contre mon modèle. L'apprentissage menant à la réactivité peut prendre plusieurs années. En effet, avant d'être un expert, c'est-à-dire avant d'être réactif, le joueur de tennis peut travailler plusieurs années.

Passons maintenant à la seconde objection. Selon la description qui vient d'être donnée, le passage à la réactivité se caractérise par un comportement qui, sur une période de temps plus ou moins longue, est sous-optimal. Comme on l'a dit, lorsque, au tout début, les intentions de bas niveau sont éliminées au profit d'une intention supérieure et que les politiques initialement contrôlées par ces intentions de bas niveau sont fusionnées, les actions* (les décisions de faire attention à une affordance) reçoivent des probabilités *arbitrairement* choisies. La plupart du temps, cela signifie que ces actions* se produisent avec des probabilités sous-optimales. Or, lorsque l'agent était un planificateur exemplaire, il exécutait ces actions* avec une probabilité optimale. En effet, les plans étaient instaurés par le GPS et celui-ci, en tant que planificateur exemplaire, introduisait l'intention de bas niveau avec une probabilité optimale. Par exemple, il produisait 1 fois sur 10 l'intention de jouer un amorti et il produisait 9 fois sur 10 l'intention de jouer en fond de terrain. Cela signifie que l'action* (la décision de faire attention à l'affordance « amortis-moi ») se produisait avec une probabilité de 10% tandis que l'autre action* (la décision de faire attention à l'affordance « frappe un coup en fond de terrain ») se produisait avec une probabilité de 90%. Donc, comme le planificateur exemplaire impliquait des actions* optimales (ou pertinentes), le passage à la réactivité se caractérise par une perte temporaire d'optimalité. Or, pourrait-on faire valoir, il n'est pas sûr que, après avoir été un planificateur exemplaire, le comportement de l'agent se caractérise pas une perte temporaire d'optimalité ou de pertinence. À ceci, je réponds qu'il s'agit là d'une question empirique. Pour trancher cette question, il faudrait examiner le comportement de l'agent qui s'achemine vers l'expertise et voir si, lorsqu'il *commence* à répondre de manière réactive aux situations fortement dynamiques (c'est-à-dire lorsqu'il commence à répondre aux bons coups de son adversaire sans être en retard), le choix de ses coups laisse à désirer. Cela n'est pas exclu. S'il s'avérait que c'était le cas, le

présent modèle serait corroboré. Par ailleurs, il faut voir que, lorsqu'il commence à devenir réactif, les coups du joueur, bien qu'ils ne soient pas optimaux, sont néanmoins meilleurs que les coups du planificateur exemplaire. En effet, face aux coups de son adversaire qui nécessitent une réaction très rapide, les coups (sous-optimaux) de ce joueur ont néanmoins l'avantage d'être exécutés à temps.

La dernière objection qui pourrait être adressée au présent modèle met en doute la pertinence de faire précéder le passage à la réactivité par un stade de planification. Au lieu de poser que, avant d'être réactif, l'agent doive se construire des plans détaillés, ne pourrait-on pas supprimer ce stade de planification et immédiatement poser l'existence d'un apprentissage qui permette à l'agent de devenir réactif? À mon avis, le stade de la planification est nécessaire. Le travail computationnel qui est exigé par la création d'une politique optimale est considérable. Il faut donc répartir cette charge computationnelle sur des instances de contrôle différentes. Ces instances de contrôle sont précisément les intentions de l'agent. Chaque intention est responsable de la création d'une politique qui s'applique à un sous-domaine du domaine total couvert par la tâche que poursuit l'agent, disons gagner un point au tennis. Une fois qu'une politique a maîtrisé un sous-domaine, qu'elle a créé une politique optimale pour ce sous-domaine, elle peut être « fusionnée » à une autre politique. Initialement, la dimension du domaine couvert par ces deux politiques était trop exigeante computationnellement pour que ce domaine puisse *directement* faire l'objet d'une politique. Mais, maintenant que ces deux politiques ont été créées, il est plus facile de créer une politique qui couvre ce domaine plus large. Pour ce faire, on procède de la manière décrite plus haut, c'est-à-dire en « fusionnant » ces deux politiques. Donc, apparemment, l'existence d'un plan (des intentions) semble bel et bien devoir précéder le passage à la réactivité.

Soulignons, pour terminer, que le passage à la réactivité peut être réalisé à différents degrés. Pour certains types d'action, et selon les capacités computationnelles de l'agent, la « fusion » des politiques peut se réaliser jusqu'à l'intention supérieure. À ce moment, l'agent sera complètement réactif. Pour d'autres types d'action, il est peu probable de pouvoir atteindre la réactivité complète puisque, compte tenu de la complexité de la tâche ou des limites computationnelles de l'agent, il n'est pas possible de « fusionner » davantage les politiques. Dans ce cas, l'agent restera partiellement planificateur.

CONCLUSION

La présente thèse a d'abord (1ère partie) tenté de définir les éléments qui font partie de l'action motrice, que ce soit celle de l'expert ou du novice. Ces éléments sont l'intention, la perception d'affordance, la perception d'une variable optique et le contrôleur moteur. Lorsque l'agent passe de novice à expert, ces éléments subissent des changements. La présente thèse a ensuite (2ème partie) tenté de définir les processus d'apprentissage qui pourraient être à la base de ces changements. Le passage à l'expertise motrice est réalisé à travers plusieurs changements. Il m'a donc fallu trouver un processus d'apprentissage pour chacun de ces changements. Je me suis très peu attardé au processus d'apprentissage qui est à la base des changements du plan, mais je me suis concentré sur les processus d'apprentissage qui sont à la base des autres changements. Le premier changement sur lequel je me suis attardé est le changement que subit le contrôleur moteur. Pour expliquer ce changement, j'ai présenté un mécanisme d'apprentissage fortement inspiré de Ashby (chapitre 7). Un autre changement est celui qui concerne la perception des affordances et la perception des variables optiques. Pour expliquer ce changement, je me suis référé à l'apprentissage par renforcement tel que présenté par Sutton et Barto (1998). C'est au cours du chapitre 8 que j'ai présenté cette forme d'apprentissage. Le dernier changement concernait l'élimination progressive du plan. Ce changement est aussi l'œuvre d'un apprentissage par renforcement (chapitre 9).

Il y a une chose dont n'a pas parlé la présente thèse et il serait important d'aborder ce sujet en conclusion. Ce thème est l'ordre d'apparition de ces changements et de leur processus d'apprentissage correspondant. Ces changements (et leur apprentissage correspondant) ne se produisent pas indépendamment les uns des autres; il existe des relations de dépendance importantes entre ces changements et entre leur apprentissage correspondant, si bien qu'un

changement (et son apprentissage correspondant) ne peut pas se produire sans qu'un autre changement (et son apprentissage correspondant) l'ait précédé. Ce sont ces relations de dépendance que j'aimerais décrire dans cette conclusion.

Tout d'abord, rappelons la structure « cognitive » qui est à la base de l'action du novice. Lorsqu'il est novice, l'agent commence par élaborer un plan hiérarchique. L'intention située au bas de ce plan hiérarchique cause une perception d'affordance. Cette perception d'affordance cause à son tour la perception d'une variable optique[85], qui agit sur le contrôleur moteur, lequel active les muscles. Lorsqu'il est novice, les quatre éléments de cette structure « cognitive » — l'intention, la perception de l'affordance, la perception de la variable optique et le contrôleur moteur — sont imparfaits. À l'aide de processus d'apprentissage, l'agent doit changer ces éléments. Comme je l'ai dit ci-dessus, les processus d'apprentissage qui permettent d'apporter ces changements ont été décrits au cours de ma thèse. Il s'agit maintenant de décrire l'ordre d'apparition de ces changements. Commençons d'abord par examiner les changements apportés au plan et voyons si ceux-ci dépendent des changements apportés aux trois autres éléments qui sont à la base de l'action motrice.

Des motifs d'ordre cognitif peuvent être à la base des changements du plan de l'agent. Par « motifs cognitifs », j'entends des connaissances que l'agent possède et qui, par le biais d'une réflexion ou d'un raisonnement, peuvent lui permettre de changer son plan. Ainsi, si l'agent veut (ou à l'intention d') aller le plus rapidement possible au cinéma et s'il sait (ou s'il a la connaissance) que d'importantes constructions ont lieu sur la rue X, il peut, via un peu de réflexion, changer son plan initial (prendre la rue X) pour un autre plan (passer par la rue Y). Lorsque l'agent change son plan sur la base de motifs cognitifs, l'agent n'a pas besoin d'attendre que des changements soient apportés aux trois autres éléments. En effet, les trois autres éléments (la perception de l'affordance, la perception de la variable optique et les paramètres de contrôle des structures coordinatives) sont responsables du mouvement de l'agent et, lorsque l'agent change son plan sur la base de motifs cognitifs, ce changement ne dépend pas de l'exécution du mouvement et des récompenses qui en découlent. Ce cas de

[85] Il peut sembler étrange de dire qu'une perception (la perception de l'affordance) cause une autre perception (la perception de la variable optique). C'est pourtant ce que le chapitre 1 a permis d'établir. Par exemple, la perception d'une cafetière cause une seconde perception, la perception qui porte sur la quantité de café qui reste dans la cafetière.

figure, c'est-à-dire la situation où l'agent change son plan sur la base de motifs cognitifs, ne m'intéresse pas. La situation qui m'intéresse est celle où le plan de l'agent est changé, non pas sur la base de motifs cognitifs, mais sur la base de récompenses. En effet, il peut arriver que ce qui motive l'agent à changer une intention soit les récompenses qu'elle procure. C'est ce cas de figure qui m'intéresse. Or, comme nous allons le voir, il est clair que, dans ce cas, l'agent doit changer les trois autres éléments avant de changer son plan.

Les récompenses que procure une intention dépendent de l'état environnemental qui découle de l'adoption de cette intention. En effet, ce sont les états environnementaux qui procurent des récompenses. Or, l'état environnemental est causé par le mouvement de l'agent. Donc, avant de pouvoir évaluer les récompenses que procure une intention et, donc, avant de pouvoir changer cette intention, il faut s'assurer au préalable que l'agent exécute le mouvement qui satisfait l'intention, c'est-à-dire le mouvement qui provoque l'état environnemental auquel réfère l'intention. Or, le mouvement est le produit des trois autres éléments de notre structure « cognitive », c'est-à-dire la perception de l'affordance, la perception de la variable optique et le contrôleur moteur. Donc, avant de changer l'intention, il faut s'assurer que ces trois éléments soient appropriés, c'est-à-dire qu'ils produisent un mouvement qui provoque l'état environnemental auquel réfère l'intention. Or, d'entrée de jeu, nous avons posé que, lorsque l'agent était novice, ces trois éléments étaient inappropriés; ils ne produisent pas un mouvement qui provoque l'état environnemental auquel réfère l'intention. Donc, avant de recevoir les récompenses qui reviennent de « droit »[86] à l'intention et, ainsi, avant de décider si, oui ou non, cette intention doit être changée, ces trois éléments doivent être changés. La question, maintenant, est de savoir lequel des trois doit être changé en premier.

Comme l'a montré la description des processus d'apprentissage qui sont à la base des changements qui touchent ces trois éléments, ces derniers sont aussi changés sur la base des

[86] Les récompenses qui reviennent de « droit » à l'intention sont les récompenses qui accompagnent l'état environnemental auquel réfère l'intention. Lorsque la perception de l'affordance, la perception de la variable optique et la valeur des paramètres de contrôle sont inappropriées, l'état environnemental qui découle de l'adoption de l'intention ne correspond pas à l'état environnemental auquel réfère cette intention. Donc, lorsque l'un ou l'autre de ces trois éléments est inapproprié, les récompenses qui reviennent à l'intention ne sont pas les récompenses qui lui reviennent de « droit ».

récompenses qu'ils procurent[87]. Et leurs récompenses correspondent aussi aux changements environnementaux qui découlent de l'exécution des mouvements. Donc, nous pouvons dire que ces trois éléments sont changés sur la base du mouvement exécuté. Si le mouvement cause un état environnemental qui donne beaucoup de récompenses, l'élément est conservé; si le mouvement cause peu de récompenses, l'élément est changé. Or, parmi ces trois éléments, certains sont plus directement responsables que d'autres de l'exécution du mouvement. En fait, les choses vont comme suit : la perception de l'affordance cause la perception de la variable optique, la variable optique agit causalement sur le contrôleur moteur et le contrôleur moteur cause le mouvement en agissant directement sur les muscles. Donc, l'élément qui est le plus directement responsable de la production du mouvement est le contrôleur moteur. Ensuite, c'est la perception de la variable optique qui est le plus directement responsable du mouvement. Enfin, l'élément qui est le plus éloigné de l'exécution du mouvement est la perception de l'affordance. Donc, apparemment, comme il est le plus directement responsable du mouvement, le contrôleur moteur semble être le premier à devoir être changé; et la perception de la variable optique et la perception de l'affordance suivraient dans l'ordre. On pourrait justifier de manière plus convaincante cet ordre d'apparition. Commençons par rendre plus convaincante l'idée que, avant de changer la perception de l'affordance, il faille changer la perception de la variable optique et la valeur qui est donnée aux paramètres de contrôle des structures coordinatives.

On change la perception de l'affordance sur la base des récompenses qui lui reviennent de « droit » et les récompenses qui lui reviennent de « droit » sont fonction du mouvement auquel elle réfère. En effet, si le mouvement produit ne correspondait pas au mouvement auquel réfère la perception de l'affordance, il ne viendrait à l'idée de personne de dire que les récompenses qui accompagnent ce mouvement sont les récompenses qui reviennent de « droit » à la perception de l'affordance. Nous pouvons donc affirmer que, avant de changer la perception de l'affordance, il faut que la perception de l'affordance cause le mouvement auquel elle réfère. Or, la perception de l'affordance cause le mouvement auquel elle réfère uniquement si la perception de la variable optique et la pondération du réseau de neurones (le

[87] Cela n'est pas tout à fait vrai en ce qui concerne l'apprentissage de la bonne pondération du réseau de neurones. Mais, pour simplifier notre description, nous pouvons supposer dans ce qui suit que c'est le cas.

contrôleur) sont appropriées. Cela découle du fait que la perception de la variable optique et le contrôleur moteur sont plus directement responsables du mouvement que la perception de l'affordance. Donc, avant de changer la perception de l'affordance, il faut que la perception de la variable optique et la pondération du réseau de neurones (le contrôleur moteur) soient appropriées. Or, nous avons posé que, initialement, la perception de la variable optique et la pondération du réseau de neurones étaient inappropriées. Donc, avant de changer la perception de l'affordance, il faut changer la perception de la variable optique et la pondération du réseau de neurones (le contrôleur); il faut les changer de manière à les rendre appropriées.

L'argument suivant, qui est de forme analogue au précédent, permet de montrer qu'il faut changer la pondération du réseau de neurones (le contrôleur) avant de changer la perception de la variable optique.

Avant de décider s'il faut changer la perception de la variable optique, il faut savoir si la perception de la variable optique cause le mouvement auquel réfère la perception de l'affordance. Si la perception de la variable optique ne cause pas ce mouvement, elle donnera un minimum de récompenses et elle ne sera pas conservée; si elle cause ce mouvement, elle donnera un maximum de récompenses et elle sera conservée. Or, la perception de la variable optique cause le mouvement auquel réfère la perception de l'affordance uniquement si la pondération du réseau de neurones (le contrôleur) est appropriée. Cela découle du fait que le contrôleur moteur est plus directement responsable du mouvement que la perception de la variable optique. Donc, avant de décider s'il faut changer la perception de la variable optique, il faut tenter de donner une pondération appropriée au réseau de neurones (le contrôleur). Or, nous avons posé que, initialement, la pondération du réseau de neurones (le contrôleur) était inappropriée. Donc, avant de décider s'il faut changer la perception de la variable optique, il faut changer la pondération du réseau de neurones (le contrôleur).

À première vue, le contrôleur moteur n'a pas besoin d'attendre qu'un autre élément change avant que la valeur de sa pondération soit changée. Cela découle du fait qu'il n'y a aucun élément qui vient s'interposer entre le contrôleur moteur et la production du mouvement.

Donc, il semble que, dans l'ordre, on doive changer 1) la pondération du réseau de neurones (le contrôleur), 2) la perception de la variable optique et 3) la perception de l'affordance. Cependant, il faut faire attention à ceci : il ne s'agit pas de compléter définitivement les changements qui doivent être apportés au contrôleur moteur et, ensuite, de compléter définitivement les changements qui doivent être apportés à la perception de la variable optique et, finalement, d'apporter les changements qui concernent la perception de l'affordance. Autrement dit, les changements à apporter à l'élément suivant ne commencent pas une fois que les changements à apporter à l'élément précédent ont été définitivement complétés. Pour voir précisément comment les choses se passent, je vais décrire la façon dont les changements se produisent.

Comme nous l'avons dit, le premier changement concerne la pondération du réseau de neurones (le contrôleur). Tout d'abord, grâce au mécanisme d'apprentissage inspiré de Ashby (chapitre 7), le système moteur tente de donner aux connexions du réseau de neurones un poids approprié. Mais, étant donné que, d'entrée de jeu, la variable optique est inappropriée (le voltigeur est novice et il ne fait pas attention à l'accélération de l'angle d'élévation de la balle, la variable optique appropriée), le mécanisme d'apprentissage est incapable d'attribuer un poids approprié au réseau de neurones. Comme on l'a montré au chapitre 7, si la variable de contrôle est inappropriée, les pondérations inappropriées vont se succéder. Il faudra donc que le système perceptuel change de perception; qu'il perçoive une autre variable optique.

Le système perceptuel, responsable de la saisie de la variable optique, va recueillir les récompenses associées à la variable optique inappropriée (la variable optique qui n'est pas l'accélération de l'angle d'élévation de la balle). Parce que cette variable de contrôle ne permet pas l'implémentation d'une pondération appropriée, cette variable optique va donner très peu de récompenses au système perceptuel. En effet, le système perceptuel reçoit beaucoup de récompenses uniquement lorsque le mouvement auquel réfère l'affordance est réalisé de manière motrice, et une pondération inappropriée ne permet pas de réaliser de manière motrice le mouvement auquel réfère l'affordance. Après avoir recueilli et mémorisé les récompenses associées à cette variable optique, le système perceptuel fera attention à une autre variable optique; autrement dit, il posera une autre action*. Une fois qu'il aura posé cette autre action*, c'est-à-dire une fois qu'il fera attention à cette autre variable optique, il

cherchera à connaître les récompenses que cette nouvelle action* procure. Pour ce faire, le système de contrôle va de nouveau tenter de trouver une pondération appropriée, cela en suivant le mécanisme d'apprentissage décrit au chapitre 7. Donc, lorsqu'il essaie une autre action* (une décision de faire attention à cette variable optique plutôt qu'aux autres), le système perceptuel doit laisser au système de contrôle le temps de trouver une valeur paramétrique appropriée, si tant est qu'une telle valeur paramétrique existe[88]. Comme nous venons de le dire, c'est pour savoir combien de récompenses la variable optique rapporte que le système perceptuel laisse au système de contrôle le temps de trouver une valeur paramétrique appropriée. On constate donc ici un va-et-vient entre le système perceptuel et le système de contrôle. Une variable optique est essayée par le système perceptuel et le système de contrôle tente de trouver une valeur paramétrique appropriée. Ensuite, une nouvelle variable optique est essayée, et le système moteur tente de nouveau de trouver une valeur paramétrique appropriée; et ce va-et-vient continue jusqu'au moment où le système perceptuel a essayé toutes les variables optiques. Comme on l'a dit, tout ce va-et-vient entre système perceptuel et système de contrôle permet au système perceptuel d'associer des récompenses à chacune des actions*. Donc, ce va-et-vient permet au système perceptuel d'évaluer sa politique (un « mapping » entre des ensembles de variables optiques et des actions*). Mais il faut noter ceci à propos de l'évaluation de la politique : parmi les actions* que le système va essayer, il s'en trouve une qui permet la réalisation motrice du mouvement « prescrit » par l'affordance. Dans notre exemple, il s'agit de l'action* qui consiste à faire attention à l'accélération de l'angle d'élévation de la balle. Lorsque cette variable optique est essayée par le système perceptuel, ce dernier cherche à connaître les récompenses qu'elle rapporte. Le système de contrôle agira donc de nouveau sur le poids des connexions du réseau de neurones (ses paramètres) et il tentera de leur trouver une valeur appropriée. Comme la variable optique est *appropriée*, le mécanisme d'apprentissage présenté au chapitre 7 (le mécanisme inspiré de Ashby) nous assure, compte tenu de sa stabilité, que le réseau de neurones sera paramétrisé de manière appropriée. Donc, lorsque le système perceptuel perçoit l'accélération de l'angle d'élévation de la balle (et après que le système moteur ait donné une valeur appropriée au poids des connexions du réseau de neurones), le

[88] Une valeur paramétrique appropriée existe uniquement si la variable optique que le système perceptuel essaie est une variable optique appropriée.

système perceptuel reçoit un maximum de récompenses. Donc, l'évaluation de la politique va donner une plus grande valeur à l'action* de faire attention à l'accélération de l'angle d'élévation de la balle. C'est cette action* que le système perceptuel va insérer dans sa politique optimale. Dorénavant, lorsque la perception de l'affordance « attrape-moi » va agir sur la région du système perceptuel qui est responsable de la perception des variables optiques, la politique activée sera celle où l'on retrouve l'action* de faire attention à l'accélération de l'angle d'élévation de la balle.

Quand l'agent a appris à percevoir la bonne variable optique et à paramétriser de manière appropriée le réseau de neurones, l'agent est alors en mesure d'exécuter le mouvement auquel réfère la perception de l'affordance. À présent, on pourrait se demander si l'apprentissage de l'exécution adéquate du mouvement, c'est-à-dire l'apprentissage de l'exécution du mouvement qui correspond au mouvement auquel réfère la perception de l'affordance, est trop long. Je ne crois pas que cet apprentissage soit trop long. D'une part, la paramétrisation du réseau de neurones se fait conformément à l'apprentissage moteur présenté au chapitre 7. Dans ce chapitre, nous avons vu que la paramétrisation adéquate du réseau de neurones correspondait à l'établissement d'un portrait de phase où les variables principales (système R, environnement et variable d'appréciation) se comportaient avec stabilité; or, le système englobant, c'est-à-dire le système T, composé du système R, de l'environnement, de la variable d'appréciation et du système S (le poids des connexions), est un système stable, si bien que l'on peut penser que l'établissement d'un portrait de phase correspondant à un comportement stable des variables principales ne devrait pas prendre trop de temps. Et quant à la sélection de la bonne variable optique de contrôle, il y a tout lieu de penser que l'apprentissage à la base de cette sélection est encore plus court. Cela découle du fait que la variable optique appropriée est la seule à avoir des récompenses positives et que le *retour* de cette variable optique est immédiat; il ne s'inscrit pas dans un épisode comprenant plus d'une action*.

Donc, après avoir trouvé la variable optique appropriée et après avoir trouvé une pondération appropriée, le mouvement auquel réfère la perception de l'affordance est réalisé de manière motrice, et cela grâce à un processus d'apprentissage qui prend un temps

raisonnable. Ceci étant dit, revenons maintenant à notre propos principal, qui est de déterminer l'ordre d'apparition des changements.

L'exécution du mouvement auquel réfère la perception de l'affordance a des conséquences sur l'environnement. Ces conséquences environnementales peuvent correspondre ou non à l'intention qui contrôle la perception de l'affordance. Lorsqu'il est novice, l'agent ne perçoit pas l'affordance appropriée, celle qui cause l'état environnemental auquel réfère l'intention. Donc, les conséquences environnementales du mouvement causé par la perception de l'affordance ne correspondent pas à l'intention. Le système perceptuel, responsable de la perception de l'affordance, reçoit donc très peu de récompenses, voire aucune récompense, lorsqu'il perçoit cette affordance. Le système perceptuel va donc entrer dans une phase exploratoire. Il va essayer d'autres affordances. Pour chacune des affordances essayées, il devra refaire toutes les étapes décrites jusqu'ici. En effet, la nouvelle affordance, celle que l'agent essaie, a besoin d'être réalisée de manière motrice et, pour cela, il faut percevoir la bonne variable optique et il faut que le réseau de neurones soit pondéré de manière approprié. Après avoir reçu les récompenses qui découlent de ces perceptions d'affordances (des actions* ou décisions de faire attention à telle affordance plutôt qu'aux autres), le système perceptuel peut évaluer sa politique (un « mapping » entre des ensembles d'affordances et des actions*). Le système perceptuel va ensuite sélectionner l'action* qui lui procure le plus de récompenses. Il s'agit évidemment de l'action* qui consiste à faire attention à l'affordance qui permet de réaliser un mouvement dont les conséquences environnementales correspondent à l'intention. Donc, lorsque le voltigeur aura l'intention d'attraper la balle, cette intention va activer au sein du système perceptuel une politique où l'on retrouve l'action* de faire attention à l'affordance « attrape-moi ».

Lorsque l'intention contrôle la bonne affordance et que le mouvement « prescrit » par l'affordance est réalisé de manière motrice par la bonne variable de contrôle et la bonne pondération, l'intention est réalisée de manière motrice d'une façon appropriée. Elle peut donc recevoir les récompenses qui lui reviennent de « droit »[89]. Si ces récompenses ne sont

[89] Les récompenses qui reviennent de « droit » à l'intention sont les récompenses qui accompagnent l'état environnemental auquel réfère l'intention. Les récompenses qui ne reviennent pas de « droit » à l'intention sont les récompenses qui, en raison d'une perception d'affordance inappropriée ou d'une

pas suffisantes, l'agent va élaborer un nouveau plan. Les nouvelles intentions devront être à leur tour réalisées de manière motrice. Donc, pour chaque intention *nouvelle*, l'agent devra reprendre toutes les étapes qui ont été décrites ci-dessus. C'est ainsi que, progressivement, l'agent élaborera des plans maximalement efficaces. Une fois qu'il disposera de plans maximalement efficaces, l'agent procèdera au dernier changement, celui qui le mène vers la réactivité. Ce dernier changement, comme je l'ai dit au chapitre 9, est le fruit d'un apprentissage par renforcement. Plus précisément, il s'agit d'éliminer les intentions de niveau inférieur en « fusionnant » les politiques qui étaient sous le contrôle de ces intentions et, ensuite, d'appliquer un apprentissage par renforcement à cette politique « fusionnée » afin de lui donner un caractère optimal. C'est ainsi que, en « fusionnant » de plus en plus les politiques et en appliquant l'apprentissage par renforcement à chacune de ces politiques « fusionnées », le contrôle des perceptions d'affordances sera pris en charge par des intentions situées toujours plus haut dans la hiérarchie. Autrement dit, la structure « cognitive » qui est responsable de l'action sera de moins en moins occupée par un plan; l'agent deviendra de plus en plus réactif.

Pour finir, j'aimerais récapituler les grandes lignes des propos qui ont été tenus au cours du présent travail et j'aimerais mettre ces propos en relation avec cette question : les actions du novice et de l'expert sont-elles *situées*?

Dans mon modèle de l'action, l'action du novice et de l'expert découlent de la même cause interne. Donc, l'action de l'expert ne fait pas intervenir un élément causal dont serait dépourvue l'action du novice. Cependant, puisque ces deux types d'action diffèrent sur plus d'un aspect, il est clair que leur cause interne n'est pas dans le même état. Quand l'agent est novice, la cause interne est dans un état qui cause une action de novice, et lorsque l'agent est expert, cette cause interne est dans un *autre* état et c'est ce changement d'état qui explique que l'action de l'expert soit une action experte. Donc, comme l'objectif de ma thèse a été de déterminer ce qui caractérise le passage de novice à expert, j'ai dû déterminer la cause interne de l'action et j'ai dû déterminer ce qui caractérise les changements qui touchent cette cause interne lorsque l'agent passe de novice à expert. Dans mon modèle, la cause interne de l'action est tripartite; trois éléments causaux concourent à la production de l'action. Le

---

perception de variable optique inappropriée ou d'une valeur inappropriée des paramètres de contrôle, accompagnent un état environnemental qui n'est pas celui auquel réfère l'intention.

premier élément est une intention; le deuxième élément, une perception d'affordance; et le troisième élément est une perception de variable optique de contrôle. Il faut compter un autre élément parmi les causes de l'action. Cet élément est l'élément le plus directement responsable du mouvement. Il s'agit du contrôleur moteur. Ma thèse a posé l'existence de changements portant sur ces quatre éléments. Le premier changement est un changement du plan, c'est-à-dire un changement des intentions. Lorsque l'agent est novice, ses plans sont imparfaits et lorsqu'il est expert, ses plans ont changé de valeur : ils sont maximalement efficaces. Le deuxième changement est un changement perceptuel. Lorsqu'il est novice, l'agent ne perçoit pas les bonnes affordances et les bonnes variables optiques de contrôle, tandis que, une fois expert, il perçoit ces propriétés de l'environnement avec justesse. Le troisième changement est un changement moteur. Lorsque l'agent est novice, le contrôleur moteur (un réseau de neurones) n'a pas la bonne pondération, tandis qu'il a une bonne pondération une fois l'agent devenu expert. Le quatrième changement, qui se produit en dernier lieu, consiste en une diminution du plan. Lorsqu'il est novice, l'agent base son action sur un plan très détaillé, et lorsqu'il est expert, sur un plan très dépouillé, pourvu de très peu d'intentions, toutes très générales. La thèse que j'ai soutenue a donc consisté à poser l'existence de ces quatre changements. Pour défendre la plausibilité de ces quatre changements (et, donc, de ma thèse), j'ai décidé de trouver les processus d'apprentissage qui sont à leur origine. Bien sûr, il n'est pas impossible que ces changements soient le fruit de processus qui ne sont pas des processus d'apprentissage. Par exemple, il se pourrait qu'ils proviennent d'un processus de maturation. Mais cette possibilité est tellement improbable que je l'ai exclue dès le départ. L'essentiel du présent travail a donc consisté à trouver un processus d'apprentissage pour ces quatre changements. Toutefois, je me suis très peu intéressé au processus d'apprentissage qui est à la base du premier changement, le changement du plan. Cela découle du fait que j'ai préféré me concentrer sur les changements (et leur processus d'apprentissage correspondant) qui mettent en scène de manière importante le système perceptuel (les perceptions d'affordances et de variables de contrôle). On pourrait m'objecter ici que le dernier changement, le changement qui consiste en une diminution du plan, n'implique pas de manière importante le système perceptuel. Cela, comme l'a montré le chapitre 9, est faux. La diminution du plan consiste en la création d'une politique *perceptuelle* plus englobante. Mais avant de présenter ces changements et leur processus

d'apprentissage correspondant, le présent travail a pris le temps de définir les différents éléments causaux de l'action. Plus précisément, nous nous sommes intéressés à la production des intentions ou plans (chapitre 2), à la possibilité des affordances (chapitres 3, 4 et 5) ainsi qu'à la nature des variables optiques de contrôle et aux contrôleurs moteurs (chapitre 6). Les chapitres suivants ont tenté de trouver un processus d'apprentissage pour les trois changements susmentionnés. Le chapitre 7 a puisé chez Ashby (1956, 1960) un processus d'apprentissage qui pourrait expliquer le changement moteur. Le chapitre 8 a trouvé un processus d'apprentissage pour le changement perceptuel. Nous avons compris que, par un apprentissage par renforcement (Sutton et Barto, 1998), il était possible de percevoir les bonnes affordances et les bonnes variables optiques de contrôle. Enfin, le chapitre 9 a introduit l'idée que la diminution du plan se produisait par la « fusion » de politiques perceptuelles particulières (c'est-à-dire des politiques contrôlées par des intentions situées au bas de la hiérarchie) en une politique englobante, qui sera contrôlée par une intention supérieure. Cette politique englobante est stochastique; elle associe plusieurs actions* à un même état environnemental et ces actions* reçoivent des probabilités. Ces probabilités, comme elles ne sont pas optimales dès le départ, doivent changer et devenir optimales. C'est un apprentissage par renforcement du type qui a été décrit au chapitre 8 qui permet d'expliquer que les probabilités attribuées aux actions* changent et deviennent optimales. Lorsque cet apprentissage a pris fin, l'agent est optimalement réactif.

Terminons cette conclusion en mettant en relation notre conception du novice et de l'expert avec cette question : Les actions du novice et de l'expert sont-elles situées?

Une action est située si la contribution de l'environnement dans la production de l'action est aussi importante, voire plus importante, que la contribution de l'esprit. Qu'en est-il du novice et de l'expert? Leur action respective est-elle située ou non? Penchons-nous d'abord sur le cas du novice.

Selon notre modèle, lorsqu'il est novice, l'agent base son action sur un plan très détaillé, c'est-à-dire un plan dont la structure hiérarchique descend très bas, et les intentions du bas de la hiérarchie contrôlent les perceptions d'affordances. Ce sont ces perceptions d'affordances (et les variables de contrôle) qui, en dernière instance, causent l'action. Donc, comme les affordances et les variables de contrôle relèvent de l'environnement, il ne fait pas de doute

que l'environnement contribue de manière importante à la production de l'action. À ce titre, nous pouvons dire que, dans notre modèle, l'action du novice possède un caractère situé. Cependant, il faut voir que la contribution de l'environnement est contrôlée par une structure cognitive très lourde. En effet, comme on vient de le dire, c'est un plan hiérarchique considérable qui, chez le novice, contrôle les perceptions d'affordances. À ce titre, la contribution de l'esprit dans la production de l'action est très importante. Le caractère situé de l'action du novice est donc limité. Mais il est clair que l'action du novice telle qu'elle a été comprise dans le présent travail est beaucoup plus située que l'action du novice telle qu'elle a été comprise par les modèles habituels de l'action novice. Dans les modèles traditionnels de l'action novice, l'intention du bas de la hiérarchie ne cause pas, via l'action de l'environnement, une perception d'affordance. Par exemple, Wolpert (Wolpert, 1997; Wolpert et Ghahramani, 2000) pose que l'intention sert d'input à un planificateur moteur, dont la fonction est de trouver une « trajectoire » déterminée au mouvement représenté de manière conceptuelle par l'intention. Ensuite, chez Wolpert, cette « trajectoire » est envoyée vers un modèle inverse, dont la fonction est de trouver une commande motrice pour chacun des points de la « trajectoire ». On le voit, dans ce modèle, les commandes motrices ne sont pas causées par une perception d'affordance et une variable de contrôle. À ce titre, ce genre de modèle de l'action novice n'est pas situé[90].

Et qu'en est-il de l'expert? Son action est-elle située? Il n'est pas difficile de voir qu'elle est fortement située; en tout cas, beaucoup plus située que l'action du novice. Chez l'expert, la structure cognitive (le plan) qui vient contrôler la perception des affordances est beaucoup moins lourde que la structure cognitive du novice. Donc, chez l'expert, la contribution de l'esprit à la production de l'action est beaucoup moins importante que la contribution de l'environnement, ce qui fait de l'action experte une action fortement située.

[90] Le modèle de Wolpert s'applique autant au novice qu'à l'expert. Dans ce modèle, ce qui distingue le novice de l'expert, c'est le fonctionnement des composantes computationnelles. Ainsi, chez le novice, le modèle inverse et le modèle prédictif fonctionnent mal, tandis qu'ils fonctionnent bien chez l'expert.

RÉFÉRENCES

Abraham, R. H., Shaw, C. D. 1992. *Dynamics – The geometry of behaviour, 2ème edition.* Redwood City, CA: Addison-Wesley.

Ashby, W.R. 1956. *An Introduction to Cybernetics*. London: Chapman & Hall Ltd.

————. 1960. *Design for a Brain.* New York: John Wiley & Sons.

Audi, R. 1973. «Intending». *Journal of Philosophy*, vol. 70, p. 387-403.

Beardsley, M. 1978. «Intending».In *Values and Morals: Essays in Honors of William Frankena, Charles Stevenson, and Richard Brandt*, sous la dir. de A. Goldman and J. Kim, p. 163-184. Dordrecht: Reidel.

Bechtel, W., Abrahamsen, A. 2002. *Connectionism and the Mind.* Oxford: Blackwell Publishers.

Bernstein, N. A. 1967. *The Co-ordination and Regulation of Movements*. Oxford: Pergamon Press.

Bertsekas, D. P. 1995. *Dynamic Programming and Optimal Control*. Belmont, Mass.: Athena Scientific.

Bertsekas, D. P., Tsitsiklis, J. N. 1996. *Neuro-Dynamic Programming*. Belmont, Mass.: Athena Scientific.

Bratman, M. 1987. *Intention, plans, and practical reason.* Cambridge, Mass.: Harvard University Press.

Bratman, M.E., Israel, D.J., Pollack, M.E. 1988. «Plans and resources-bounded practical reasoning». *Computational Intelligence*, vol.4, no.3, p. 349-355.

Brooks, R. A. 1999. *Cambrian Intelligence.* Cambridge, Mass.: MIT Press.

Chapman, S. 1968. «Catching a baseball». *American Journal of Physics*, vol. 36, p. 368-370.

Chemero, T. 2003. «Radical empiricism through the age». *Contemporary Psychology*, vol. 48, p. 18-20.

Churchland, P. 1970. «The Logical Character of Actions-Explanations». *The Philosophical Review*, vol. 79, p. 214-236.

Cummins, R. 1989. *Meaning and Mental Representation.* Cambridge, Mass.: MIT Press.

Davis, W. 1984. «A Causal Theory of Intending». *American Philosophical Quarterly*, vol. 21, p. 43-54.

Dretske, F. 1981. *Knowledge and the flow of information.* Cambridge, Mass.: MIT Press.

———. 1988. *Explaining Behavior*. Cambridge, Mass.: MIT Press.

Dreyfus, S. E., Law, A. M. 1977. *The Art and Theory of Dynamic Programming*. New York: Academic Press.

Elman, J. L. 1990. «Finding Structure in Time». *Cognitive Science*, vol. 14, p. 179-211.

Ernst, G., Newell, A. 1969. *GPS: A Case Study in Generality and Problem Solving*. New York: Academic Press.

Fajen, B. R. 2007. «Affordance-Based Control of Visually guided Action». *Ecological Psychology*, vol. 19, no 4, p. 383-410.

———. 2005. «The sclaling of information to action in visually guided braking». *Journal of experimental psychology: Human perception and performance*, vol. 31, no 5, p. 1107-1123.

———. 2001. «Steering toward a goal by equalizing taus». *Journal of Experimental Psychology: Human Perception and Performance*, vol. 27, no 4, p. 953-968.

Fajen, B.R., Riley, M.A., Turvey, M.T. 2009. «Information, affordances, and the control of action in sport». *International journal of sport psychology*, vol. 40, no 1, p. 79-107.

Fajen, B. R., Warren, W. H. 2004. «Visual guidance of intercepting a moving target on foot». *Perception*, vol. 33, p. 689-675.

Feldman, A.G. 1966. «Functional tuning of the nervous system with control of movement or maintenance of a steady posture –II. Controllable parameters of the muscles». *Biophysics*, vol. 11, p. 565-578.

Fikes, R.E, Nilsson, N.J. « Strips: A new approach to the application of theorem proving to problem solving ». *Artificial Intelligence*, vol. 2, no 3-4, p. 189-208

Fodor, J. 1983. *Representations:Philosophical essays on the foundations of cognitive science*. Cambridge, Mass.: MIT Press.

Fodor, J., Pylyshyn, Z. 1981. «How direct is visual perception? Some reflections on Gibson's «ecological approach»». *Cognition*, vol. 9, no 2, p.139-196.

Frith, C.D., Blakemore, S-J., Wolpert, D.M. 2000. «Abnormalities in the awareness and control of action». *Philosophical Transactions: Biological Sciences*, vol. 355, no 1404, p. 1771-1788.

Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin Company.

———. 1977. «The theory of affordances». In *Perceiving, acting, and knowing: Toward ecological psychology*, sous la dir. de R.E. Shaw et J. Bransford. Hillsdale. NJ: Lawrence Erlbaum Associates.

———. 1979/1986. *The Ecological Approach to Visual Perception*. London: Lawrence Erlbaum Associates.

Giunti, Marco. 1995. «Dynamical Models of Cognition». In *Mind as motion*, sous la dir. de R.F. Port et T. van Gelder. Cambridge, Mass.: MIT Press.

Greeno, J. G. 1994. «Gibson's affordances». *Psychological Review*, vol. 101, p. 336-342.

Haruno, M., Wolpert, D. M., Kawato, M. 2001. «MOSAIC Model for Sensorimotor Learning and Control». *Neural Computation*, vol. 13, p. 2201-2220.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press

Heft, H. 1989. «Affordances and the body: An international analysis of Gibson's ecological approach to visual perception». *Journal for the Theory of Social Behavior*, vol. 19, p. 1-30.

Hess, E. 1975. The Role of Pupil Size in Communication. Réédité in *Mind and Behavior, Readings from Scientific American* (1980), sous la dir. de Atkinson et Atkinson. San Francisco: W. H. Freeman and Co.

Jordan, M. I. 1986. «Attractor dynamics and parallelism in a connectionist sequential machine». *Proccedings of the Eighth Annual Conference of the Cognitive Science Society*, p.531-546. Hillsdale, N.J : Lawrence Erlbaum.

Jordan, M. I., Wolpert, D. M. 1999. «Computational motor control». In *The Cognitive Neurosciences*, sous la dir de M. Gazzaniga. Cambridge, MA: MIT Press.

Kalos, M. H., and Whitlock, P. A. 1986. *Monte Carlo Methods*. New York: Wiley.

Kim, N. G., Turvey, M. T. 1999. «Eye movements and a rule for perceiving direction of heading». *Ecological Psychology*, vol. 11, p. 223-248.

Latash, M. L. 1996. «The Bernstein Problem : How does the Central Nervous System Make its Choices?». In *Dexterity and its Development*, sous la dir. de M.L. Latash et M.T. Turvey, p.277-304. Mahwah, New Jersey: Lawrence Erlbaum Associates

Lee, D. N. 1976. «A theory of visual control of bracking based on information about time-to-collision». *Perception*, vol. 5, p. 437-459.

Lombardo, T. J. 1987. *The reciprocity of perveiver and the environment*. Mahwah, NJ : Lawrence Erlbaum Associates.

McBeath, M. K., Shaffer, D. M., Kaiser, M. K. 1995. «How baseball outfielders determine where to run to catch a fly ball». *Science*, vol. 268, no 5210, p. 569-573.

Merleau-Ponty, M. 2002/1942. *La Structure du Comportement*. Paris : Quadrige/PUF.

Miall, R.C., Wolpert, D.M. 1996. «Forward Models for Physiological Motor Control». *Neural Networks*, vol. 9, no 8, p. 1265-1279.

Miall, R.C., Weir, D.J., Wolpert, D.M., Stein, J.F. 1993. «Is the cerebellum a Smith predictor». *Journal of Motor Behavior*, vol. 25, no 3, p. 203-216.

Millikan, R.G. 1984. *Language, Thought, and other Biological Categories*. Cambridge, Mass.: MIT Press.

Neuberg, M. 1991. *Théorie de l'action*. Liège : P. Mardaga

Newell, A., Shaw, J.C., Simon, H.A. 1957. *Preliminary description of general problem solving program-1 (GPS-I)*. CIP Work. Pap. 7. Pittsburgh: Carnegie Institute of Technology.

Peper, L., Bootsma, R. L., Mestre, D. R., Bakker, F. C. 1994. «Catching balls: How to get the hand to the right place at the right time». *Journal of Experimental Psychology: Human Perception and Performance*, vol. 2o, no 3, p. 591-612.

Reed, E. S. 1996. *Encountering the world.* New York: Oxford University Press.

Rubinstein, R. Y. 1981. *Simulation and the Monte Carlo Method.* New York: Wiley.

Ross, S. 1983. *Introduction to Stochastic Dynamic Programming*. New York: Academic Press.

Sanders, J. T. 1997. «An ontology of affordance». *Ecological Psychology*, vol. 9, p. 97-112.

Scarantino, A. 2003. «Affordances explained». *Philosophy of Science*, vol. 70, no 5, p. 949-961.

Searle, J. 1983. *Intentionality*. Cambridge, Cambridge University Press.

Simon, H. 1981. *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.

Stoffregen, T. A. 2003. «Affordances as Properties of the Animal-Environment System». *Ecological Psychology*, vol. 15, no 2, p. 115-134.

Suchman, L. A. 2007. *Human-Machine Reconfigurations. Plans and situated actions*. Cambridge: Cambridge University Press

Sutton, R. S. 1988. «Learning to predict by the method of temporal differences». *Machine Learning*, vol. 3, p. 9-44.

————. 1996. «Generalization in reinforcement learning: Successful exemples using sparse coarse coding». In *Advances in Neural Information Processing Systems: Proceeding of the 1995 Conference*, sous la dir. de D.S. Touretsky, M.C. Mozer et M.E. Hasselmo, p. 1038-1044. Cambridge: MIT Press.

Sutton, R.S., Barto, A.G. 1998. *Reinforcement Learning*. Cambridge, Mass: MIT Press.

Tuller, B., Fitch, H.L., Turvey, M.T. 1982. «The Bernstein Perspective: II. The Concept of Muscle Linkage or Coordinative Structure». In *Human Motor Behavior*, sous la dir. de J.A.S. Kelso, p. 253-270. Hillsdale, New Jersey: The Erlbaum associates.

Turvey, M.T. 1990. «Coordination». *American Psychologist*, vol. 45, p. 938-953.

Turvey, M. T. 1992. «Affordances and prospective control: An outline of the ontology». *Ecological Psychology*, vol. 4, p. 173-187.

Turvey, M.T., Shaw, R.E., Reed, E.S., Mace, W.M. 1981. «Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn». *Cognition*, vol. 9, no 3, p. 237-304.

Turvey, M.T., Fitch, H.L., Tuller, B. 1982. «The Bernstein Perspective: I. The Problem of degrees of Freedom ans Context-Conditionned Variability». In *Human Motor Behavior*, sous la dir. de J.A. Kelso, p. 239-251. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Turvey, M.T., Carello, C., Nam-Gyoon Kim. 1990. «Links between Active Perception and the Control of Action». In *Synergetics of Cognition*, sous la dir. de H. Haken et M. Stadtler, p. 269-295. Berlin: Springer-Verlag.

Van Gelder, T., Port, R. F. 1995. «It's About Time: An Overview of the Dynamical Approach to Cognition». In *Mind as Motion*, sous la dir. de T. van Gelder et R.F. Port, p. 1-43. Cambridge, Mass.: MIT Press

Wann, J. P., Land, M. 2000. «Steering with or without the flow: Is the retrieval of heading necessary?». *Trends in Cognitive Sciences*, vol. 4, no 8, p. 319-324.

Warren, W. H. 1988. «Action Modes and Laws of Control». In *Complex Movement Behavior: "The" Motor-Action Controversy*, sous la dir. de O.G. Meijer et K. Roth, p. 339-380. Amsterdam: North Holland.

————. 1984. «Perceiving affordances: Visual guidance of stair climbing». *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, p. 683-703.

Warren, W. H., Fajen, B. R. 2004. «From Optic Flow to Law of Control». In *Optic Flow and Beyond*, sous la dir. de L.M. Vaiva, S.A. Beardsley et SK. Rushton, p. 307-337. Dordrecht: Kluwer Academic Publishers.

White, D. J. 1969. *Dynamic Programming*. San Francisco: Holden-Day.

Whittle, P. 1982. *Optimization over Time, vol. 1*. New York: Wiley.

————. 1983. *Optimization over Time, vol. 2*. New York: Wiley.

Wolpert, D.M. 1997. «Computational approaches to motor control». *Trends in Cognitive Sciences*, vol. 1, no 6, p. 209-216.

Wolpert, D.M., Ghahramani, Z. 2000. «Computational principles of movement neuroscience». *Nature Neuroscience*, vol. 3, p. 1212-1217.

Yilma, E. H., Warren, W. H. 1995. «Visual control of braking: a test of the tau hypothesis». *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no 5, p. 996-1014.

Zednik, C. 2011. «The Nature of Dynamical Explanation». *Philosophy of Science*, vol. 78, no 2, p. 238-263.