

UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS VI
ÉCOLE DOCTORALE 3C
Cerveau - Cognition - Comportement

THÈSE

pour obtenir le titre de

Docteur en Sciences

Mention : SCIENCES COGNITIVES

Présentée et soutenue par

Lionel RIGOUX

**Compromis entre efforts
et récompenses :
Un modèle unifié de la décision
et de la motricité**

*Thèse dirigée par Emmanuel GUIGON
et préparée à*

l'Institut des Systèmes Intelligents et de Robotique
CNRS UMR 7222

soutenue le 24/11/2011

Jury :

<i>Rapporteurs :</i>	Étienne BURDET	Professeur, Imperial College London
	Emmanuel PROCYK	Chargé de Recherche, CNRS
<i>Examineurs :</i>	Thomas BORAUD	Directeur de Recherche, CNRS
	Jean-René DUHAMEL	Directeur de Recherche, CNRS
	Emmanuel GUIGON	Chargé de Recherche, CNRS
	Mathias PESSIGLIONE	Chargé de Recherche, INSERM
	Michèle THIEULLEN	Maître de Conférences, UPMC

Résumé :

Le comportement est fortement déterminé par les coûts (*e.g.* les efforts) et les bénéfices (*e.g.* la nourriture, l'argent) de nos actions. Dans les approches actuelles, en économie et en écologie plus particulièrement, l'optimisation de ces valeurs sert de principe pour organiser le comportement. Toutefois, ces approches ne peuvent expliquer comment la décision (le choix d'un objectif à atteindre) est effectivement convertie en action (le contrôle de la bio-mécanique du corps). La raison à cela est que l'optimisation des coûts et des bénéfices sont considérés comme deux processus imperméables l'un à l'autre. Pourtant, ces deux éléments interagissent fortement à tous les niveaux de l'élaboration de nos actes.

Ce travail propose un modèle du comportement dans lequel la décision et la production motrice émergent de l'optimisation permanente du compromis prospectif et pondéré entre la récompense et l'effort moteur liés à l'action. Ce modèle permet de rendre compte de la décision dans les situations d'effort ainsi que des caractéristiques détaillées de la coordination motrice et de sa modulation par le contexte comportemental. Le formalisme présenté offre un cadre normatif pour interpréter les bases neurologiques de la décision et du contrôle moteur. Nous proposons également des éléments théoriques pour comprendre le rôle de la dopamine et des ganglions de la base dans la régulation de l'effort, plus particulièrement dans les pathologies telles que la maladie de Parkinson.

Mots clés : modèle, contrôle moteur, décision, motivation, effort, dopamine

Remerciements

Je tiens tout d'abord à remercier Emmanuel Guigon pour m'avoir offert son savoir et son temps durant ces trois années. Sa passion et sa rigueur m'ont guidé à chaque instant, et je lui suis profondément reconnaissant de m'avoir fait profiter de sa finesse et de sa connaissance illimitée de la littérature au gré de nos échanges toujours sympathiques. Je souhaite également remercier Olivier Sigaud pour avoir partagé ses idées et sa vivacité lors de nos trop peu nombreuses interactions. Son optimisme et sa curiosité m'ont stimulé tout au long de ma thèse. Je tiens également à exprimer toute ma reconnaissance à Alexander Terekhov pour m'avoir initié, avec patience, aux arcanes de l'optimalité, et pour toutes ces discussions qui ont scellé notre amitié. Je remercie également Serena Ivaldi pour m'avoir démontré qu'aucune dérivation ne saurait résister à un crayon à papier ; ses calculs m'ont fait gagner en confiance et un temps précieux. Je salue aussi le groupe du bureau J12 ; Ildar, Chao, Amir, Mohamed, Michael, Rafal, Jeremy, Hiba, Carlo, Jonathan et tous ceux qui ont donné de la chaleur à mon séjour à l'ISIR, avec une pensée spéciale pour Bruno et Stéphane. Je suis également reconnaissant à toute ma famille pour son soutien permanent et indéfectible, et pour m'avoir donné la chance de faire cette thèse. Je souhaite également remercier Flo pour tout ce que nous avons partagé, pour sa confiance et ses encouragements tout au long de cette aventure, et pour avoir été (et pour longtemps encore j'espère), au gré des vents, toujours à mes côtés.

Table des matières

1	Introduction	1
2	Modéliser le comportement	7
2.1	Optimiser le bénéfice de son comportement	7
2.1.1	Notion de récompense	7
2.1.2	Anticiper le bénéfice de ses actions	8
2.1.3	Le principe de maximum de récompense	9
2.2	La question de l'effort	10
2.2.1	La coordination du mouvement	12
2.2.2	Le principe du moindre effort	13
2.3	Effort et durée de mouvement	15
2.3.1	La régulation du temps de mouvement	16
2.3.2	Le coût du temps	18
2.4	Flexibilité du mouvement et du comportement	20
2.4.1	L'architecture du contrôle en ligne	21
2.4.2	Emergence de l'action	23
2.5	Un modèle général du comportement	24
2.5.1	Notion de dynamique	24
2.5.2	Le principe d'optimalité	26
2.5.3	Sélection d'une durée de mouvement optimale	29
2.5.4	Exécution en ligne du mouvement	32
3	Méthodologie	35
3.1	Optimisation du contrôleur	35
3.1.1	Résolution du contrôle optimal	35
3.1.2	Détermination du temps optimal	40
3.2	Architecture générale du contrôle	41
3.2.1	La boucle de contrôle	42
3.2.2	Déroulement de la simulation d'une action motrice	44
3.3	Modélisation de l'objet contrôlé	50
3.3.1	Objet I : point masse	50
3.3.2	Objet II : point masse avec muscle	50
3.3.3	Objet III : bras articulé	52
3.4	Nature des paramètres	59
3.4.1	Paramètres de la tâche	60
3.4.2	Paramètres internes	60
4	Le contrôle du mouvement	63
4.1	Propriétés fondamentales du mouvement	63
4.1.1	Electro-myogramme	64
4.1.2	Formation des trajectoires	65

4.2	Effort et temps de mouvement	66
4.2.1	Loi de mise à l'échelle	67
4.2.2	Anisotropies inertielles	68
4.3	Statistiques du mouvement	69
4.3.1	Profil de variabilité	70
4.3.2	Loi de Fitts	71
4.4	Flexibilité du mouvement face aux perturbations	74
4.4.1	Déplacements de la cible	75
4.4.2	Altération de la dynamique du corps	77
5	L'organisation du comportement	83
5.1	Effort et décision	84
5.1.1	Procédure de décision	84
5.1.2	Situation de conflit entre effort et récompense	85
5.1.3	Une représentation commune de l'effort	88
5.2	Sensibilité du mouvement au contexte comportemental	91
5.2.1	Motivation et vitesse	91
5.2.2	Le cas des séquences motrices	92
6	Modulation de l'action et pathologies	97
6.1	Effet de la vigueur	98
6.2	Effet du décompte	99
6.3	Deux modes de régulation	100
6.3.1	L'hypothèse dopaminergique	100
6.3.2	Le rôle du décompte	102
7	Discussion générale	105
7.1	Un comportement moteur flexible et unifié	105
7.2	Une redéfinition de la frontière entre décision et mouvement	108
7.2.1	Conséquences pour l'architecture du contrôle du comportement	110
7.2.2	Le paramètre de décompte	113
7.3	Conclusion	115

Introduction

Imaginons un singe se déplaçant d'arbre en arbre dans une jungle du Nouveau Monde. Sa survie dépend tout d'abord de sa capacité à choisir de façon avisée à quelles branches s'agripper afin d'y cueillir les fruits qui le nourriront ou y trouver la compagnie de ses pairs. Il doit également faire preuve d'une grande agilité pour arriver à se mouvoir au travers des ramifications complexes qui l'entourent et ainsi atteindre les ressources qu'il convoite sans toutefois gaspiller ses précieuses forces (Stephens et Krebs, 1986).

Cet exemple illustre le fait que nous devons constamment décider de quelle manière utiliser notre corps afin de tirer le meilleur parti de notre environnement. Comme nous ne pouvons pas tout faire à la fois, nous sommes en permanence obligés de choisir, parmi toutes les opportunités d'agir qui se présentent, celles qui nous paraissent les plus avantageuses. Nous devons de plus éviter de dilapider notre énergie dans des actions inutiles ou peu rentables, sous peine de nous épuiser sans jamais retirer de bénéfices du monde qui nous entoure.

L'approche classique pour aborder les problèmes de décision prend la forme de modèles normatifs, c'est-à-dire qui cherchent à identifier les lois à l'origine du comportement. Ces modèles reposent sur la notion d'optimalité, qui fournit une traduction naturelle des concepts de coûts et de bénéfices et permet surtout d'expliquer comment ces valeurs peuvent contribuer de façon opératoire à l'élaboration du comportement. Selon cette approche, l'action peut être expliquée par l'intervention de deux méca-

nismes séparés : la décision et la production du mouvement (Rangel et al., 2008). Ces processus permettraient successivement de maximiser nos bénéfices (en décidant *où* aller) et de minimiser nos efforts (en décidant *comment* y aller).

D'un côté, la décision serait le processus par lequel les animaux sélectionnent dans l'environnement les ressources qui semblent les plus profitables et vers lesquelles diriger leurs actes (Wakker et Kahneman, 1997; Glimcher et al., 2005; Sugrue et al., 2005). Dans notre exemple, c'est à ce moment que notre singe établira qu'il préfère aller récupérer tel fruit plutôt que tel autre moins mûr, aller manger plutôt que de rejoindre son groupe, *etc.*

Le processus de contrôle du mouvement est le mécanisme qui permet à notre singe d'utiliser efficacement ses muscles pour se déplacer vers la destination qu'il a choisie. Les théories actuelles du mouvement postulent que la production motrice repose quant à elle sur un principe de minimum de coût (*e.g.* minimum d'effort) indépendant de la valeur comportementale de l'action (Wolpert et Ghahramani, 2000; Todorov et Jordan, 2002; Krakauer et Shadmehr, 2007; Körding, 2007).

Il apparaît cependant clairement que les modèles normatifs actuels ne sont pas en mesure de rendre compte du comportement de notre singe. La principale raison à cela est que les processus de décision et d'exécution de l'action sont considérés séparément, comme s'ils étaient imperméables, aveugles l'un à l'autre. Si nous compliquons un peu la situation, il apparaît pourtant que la décision et la motricité sont en réalité étroitement intriquées et qu'une séparation stricte de ces mécanismes ne permet pas d'expliquer la complexité du comportement moteur.

Tout d'abord, le coût associé à la réalisation des actions influence fortement la décision (Stephens et Krebs, 1986; Denk et al., 2005; Floresco et al., 2008; Croxson et al., 2009; Prévost et al., 2010; Kurniawan et al., 2010; Cos et al., 2011). Si par exemple notre singe a le choix entre un arbre assez éloigné et chargé de fruits et un

arbre plus proche mais moins garni, l'effort et le temps que nécessiterait le déplacement vers l'arbre le plus rempli risque de dissuader notre animal et de faire pencher sa décision vers la solution de facilité.

Aussi, les paramètres du mouvement sont modulés par de nombreux facteurs internes et environnementaux (Montagnini et Chelazzi, 2005; Aarts et al., 2008; Xu-Wilson et al., 2009). En particulier, la valeur de l'objectif visé n'influence pas seulement la décision mais également la vitesse d'exécution des actions (Takikawa et al., 2002). Si par exemple notre singe a particulièrement faim et s'il ne veut pas prendre le risque qu'un de ses congénères ne lui passe devant, il aura tendance à accélérer ses mouvements, et donc à dépenser plus d'effort, pour attraper la nourriture qu'il vise.

Ainsi, il est impossible d'expliquer nos choix sans incorporer le coût de réalisation des actions. De même, il est impossible d'expliquer nos mouvements sans expliciter leur perméabilité au contexte comportemental. Limiter la décision à la maximisation des récompenses et le contrôle moteur à la minimisation des efforts ne permet donc pas de rendre compte du comportement de façon générale. En particulier, les théories actuelles n'expliquent pas comment l'effort est régulé et intégré dans l'élaboration du comportement.

L'enjeu pour la survie que représentent la préservation de nos ressources et l'exploitation de celles de l'environnement semble un déterminant majeur de l'organisation du comportement animal (Stephens et Krebs, 1986). L'élaboration de l'action apparaît ainsi comme la réponse économique à un problème écologique, celui de l'exploitation judicieuse du monde qui nous entoure. Un des objectifs de cette thèse est d'identifier les critères qui guident ainsi le choix de nos actes, c'est-à-dire de comprendre quelles règles nous poussent à traduire un besoin en une action plutôt qu'en une autre.

Nous commencerons dans le chapitre 2 par dresser une revue des données empiriques et des modèles normatifs de la décision et du mouvement avec pour objectif de construire un cadre formel capable d'expliquer la traduction de nos objectifs en actes moteurs. Nous utiliserons pour cela le langage de l'optimalité afin d'élaborer un modèle normatif du comportement. Notre travail se fonde sur un modèle computationnel qui exprime dans un cadre unique les processus de décision et de coordination : à chaque instant d'un comportement, la bonne décision à prendre ainsi que la bonne façon de diriger le corps sont définies par la maximisation d'un compromis à long terme entre coûts et bénéfices. La formulation du modèle s'inscrit dans le cadre de la théorie du renforcement et du contrôle optimal (Doya, 2000; Todorov et Jordan, 2002), au carrefour entre l'écologie (Stephens et Krebs, 1986) et l'économie (Sethi et Thompson, 1972).

Nous décrivons ensuite, dans le chapitre 3, une méthode pour simuler notre modèle dans des tâches réalistes de décision et de mouvement. Grâce à cette approche computationnelle, nous serons capables d'analyser dans quelle mesure les principes proposés peuvent effectivement expliquer les propriétés du comportement. En particulier, nous voulons établir les liens entre décision et motricité et explorer les mécanismes de régulation qui peuvent les sous-tendre. La validation de ce modèle repose sur la reproduction d'expériences issues de la littérature sur la décision et le contrôle moteur. La confrontation qualitative et quantitative avec les données expérimentales offre un outil d'interprétation et de compréhension unique car fondé sur un processus génératif plutôt que descriptif.

Dans le chapitre 4, nous testerons la capacité de notre modèle à reproduire les propriétés caractéristiques du mouvement et ainsi à expliquer le comportement dans ses aspects moteurs. Nous établirons ensuite dans le chapitre 5 comment notre formalisme permet d'aborder les influences réciproques entre contexte comportemental et production motrice, et plus particulièrement la décision en situation d'effort et la

formation des séquences motrices. Enfin, nous analyserons la signification des paramètres du modèle pour tenter de comprendre les mécanismes en jeu dans la régulation du comportement.

Nous terminerons par une discussion générale de nos résultats (chapitre 7), en mettant l'accent sur leur intégration dans une théorie plus globale de la neuro-biologie de l'action.

Modéliser le comportement

2.1 Optimiser le bénéfice de son comportement

Nous cherchons à comprendre comment un animal élabore ses actions en fonction de ses besoins et des situations qu'il rencontre dans son environnement. Afin d'aborder ce problème de façon formelle, nous allons nous placer dans le cadre très général de la *théorie de l'apprentissage par renforcement* (Sutton et Barto, 1998; Bertsekas, 1995). Ce formalisme permet de capturer la manière dont un animal utilise sa représentation du monde pour interagir avec son environnement et atteindre ses buts.

Commençons par modéliser formellement notre créature. Nous parlerons de l'*état* d'un animal pour désigner de façon générale sa position, sa vitesse, ou tout autre caractéristique physique pertinente pour la description de son comportement. Pour l'instant, nous pouvons nous contenter de considérer sa position dans l'environnement. Son comportement peut alors être décrit de façon minimale par ses déplacements successifs, c'est-à-dire la séquence des états qu'il traverse.

Dans cette section, nous allons voir comment nous pouvons construire, à l'aide de la notion de récompense, un premier modèle pour capturer l'idée que l'action est un moyen de maximiser le bénéfice de son comportement.

2.1.1 Notion de récompense

Parmi toutes les sensations qu'un animal peut expérimenter lors de ses interactions avec son environnement, certaines lui apparaissent comme naturellement plaisantes,

le goût sucré d'un fruit par exemple (Berridge et al., 2009). Nous pouvons facilement imaginer que la spécificité de ces plaisirs, issue de la sélection naturelle, possède un véritable rôle opérationnel dans la survie de l'animal. Le goût d'un aliment par exemple est un relativement bon indicateur de sa qualité nutritive et préférer le sucré à l'amer procure un avantage indéniable pour la survie. Cette signalisation dépend de plus de l'état interne et des besoins immédiats de l'animal : une nourriture est bien moins plaisante une fois rassasié et une boisson sera d'autant plus appréciée que l'animal a soif. En d'autres termes, une stimulation est plus ou moins enviable selon les dispositions de l'animal (Berridge et Robinson, 2003). Ainsi, la signalisation hédonique permet de distinguer dans l'environnement les ressources utiles, les nourritures à ingérer, les événements favorables qu'il faut rechercher pour survivre (Berridge et Aldridge, 2008). D'une façon générale, nous désignerons par *récompense* la valeur positive qu'une créature attribue à un objet ou une stimulation particulière.

2.1.2 Anticiper le bénéfice de ses actions

Pour pouvoir planifier un comportement, reconnaître les choses que l'on aime n'est pas suffisant : il faut également savoir où et quand les trouver. Toutefois, grâce à son appareil sensoriel, un animal reçoit à tout instant un faisceau d'informations qui peuvent lui servir à estimer l'état du monde. Certains de ces états sont généralement annonciateurs de la disponibilité de ressources intéressantes, comme un bruit peut signaler la présence d'une proie, ou l'aspect d'un arbre indiquer l'existence de fruits par exemple. De nombreux travaux ont mis en évidence que les animaux sont ainsi capables d'anticiper les récompenses qu'ils peuvent obtenir de leurs actions à partir des informations récoltées dans l'environnement (Wise, 2004; Dayan et al., 2006; Daw et Shohamy, 2008; Glimcher, 2011).

L'anticipation des récompenses permet de mesurer l'intérêt des états du monde qu'il est possible de visiter. De plus, le cerveau semble représenter les bénéfices po-

tentiels des actions dans une « devise standard », indépendante de la nature de la stimulation récompensante (Montague et Berns, 2002). Il s'agit donc d'une mesure universelle qui ne fait aucune distinction entre les types de récompenses qui sont anticipées. Elle permet ainsi de comparer sur un pied d'égalité les différents objectifs que peuvent viser nos actions.

2.1.3 Le principe de maximum de récompense

Si notre singe a faim, il a tout intérêt à utiliser sa connaissance de l'environnement pour se diriger à l'endroit où il pense trouver de la nourriture. La capacité à prédire la valeur des récompenses disponibles dans l'environnement, en nous permettant d'anticiper le bénéfice de nos actions, peut ainsi guider le choix de nos actes moteurs (figure 2.1).

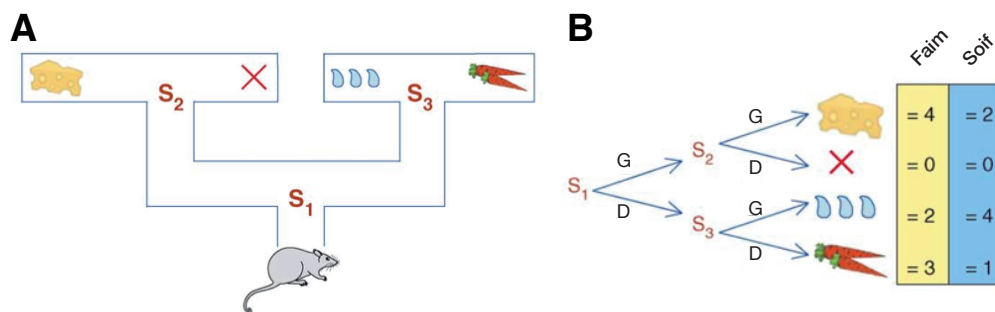


FIGURE 2.1 – *Illustration du problème de décision (adapté de Niv et al., 2006).*
A Une souris doit choisir, à plusieurs intersections, entre aller à gauche (G) ou à droite (D). La récompense qu'elle peut retirer de ses actions dépend de la succession de choix qu'elle fait. **B** En étant capable de prédire la localisation et la valeur de chaque récompense (tableau, valeurs données dans une unité arbitraire), l'animal peut optimiser son comportement et choisir d'effectuer la séquence de choix G-G s'il a faim, ou D-G s'il a soif.

L'hypothèse selon laquelle nous cherchons à maximiser le bénéfice que nous espérons tirer du monde qui nous entoure permet d'expliquer le comportement humain et animal dans de nombreuses situations de choix entre des actions discrètes (Wakker et Kahneman, 1997; Sutton et Barto, 1998; Glimcher et al., 2005; Daw et Doya, 2006;

Berridge et Aldridge, 2008). De nombreuses données neurophysiologiques, démontrant le rôle prépondérant du mécanisme de prédiction des récompenses dans la décision, confortent cette idée que le comportement repose sur un critère d'optimisation des bénéfices (Platt et Glimcher, 1999; Sugrue et al., 2005; Rangel et al., 2008).

La théorie du maximum de récompense offre ainsi un cadre normatif puissant pour expliquer les comportements orientés vers des buts. Elle permet en effet d'expliquer pourquoi nous décidons d'agir comme nous le faisons face à des situations de choix entre plusieurs récompenses.

2.2 La question de l'effort

Nous avons jusqu'ici considéré qu'une action était évaluée uniquement en regard de la récompense qu'elle permet d'obtenir. Le principe du maximum de récompense permet de capturer clairement le fait que nous organisons nos choix en fonction des bénéfices que nous espérons obtenir de nos actions. Cependant, ce formalisme néglige un facteur important de la décision : l'effort que demande la réalisation de l'action. Nous ne nous sommes en effet intéressés qu'au résultat des actions, c'est-à-dire aux états récompensés que les actes permettent d'atteindre. Bouger a pourtant un coût, et ce coût a un impact non négligeable sur nos choix.

L'équipe de Salamone a largement contribué à démontrer l'impact de la difficulté d'agir sur la décision (Cousins et Salamone, 1994; Salamone et Correa, 2002; Salamone et al., 2005, 2007, 2009) en confrontant des rats à des situations de conflit entre effort et récompense. Un des paradigmes que les auteurs ont utilisé est schématisé sur la figure 2.2.A. Un rat est placé dans un labyrinthe en T où il doit résoudre l'alternative suivante : d'un côté se trouve une petite récompense facilement accessible; dans l'autre branche, l'animal a la possibilité d'escalader une barrière pour obtenir une récompense plus importante. Si la deuxième option est la plus intéressante en terme

de bénéfiques, elle nécessite également une plus grande dépense énergétique à cause de l'obstacle. Les résultats comportementaux démontrent de façon robuste que les rats se détournent de la grande récompense si l'effort requis pour escalader la barrière est trop grand et préfèrent alors se déplacer vers la petite récompense. Différents protocoles ont permis de répliquer ce phénomène (Rudebeck et al., 2006; Walton et al., 2006), que l'effort à surmonter prenne la forme d'une répétition d'appuis sur un levier (Walton et al., 2006; Salamone et al., 2007, figure 2.2.B) ou d'une distance plus ou moins grande à parcourir (Stevens et al., 2005; Mühlhoff et al., 2011). Les études chez l'humain ont confirmé l'importance de l'effort lié à l'exécution de l'action dans la décision (Crosson et al., 2009). L'expérience de Cos et al. (2011) a de plus montré que l'évaluation de l'effort reposait sur une représentation fine du mouvement dans toute sa complexité mécanique.

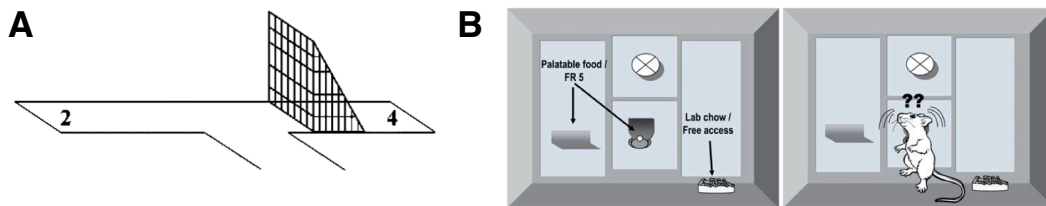


FIGURE 2.2 – *Exemples de décision en situation d'effort.* **A** Un rat est placé à l'entrée d'un labyrinthe en T. Il doit choisir entre aller dans la branche de gauche, où une petite récompense est facilement accessible, et escalader une barrière dans la branche de droite pour obtenir une récompense plus importante (Denk et al., 2005). **B** Un rat est ici placé dans une chambre de test. S'il peut accéder librement à une nourriture qu'il apprécie peu, il peut également choisir d'appuyer sur un levier pour obtenir une nourriture plus savoureuse. Si cette dernière option est plus intéressante, elle demande aussi un effort plus important (Salamone et al., 2007).

Si nous voulons expliquer le comportement, nous devons donc intégrer le coût des actions lors de leur évaluation. Le problème est que le formalisme que nous avons utilisé jusqu'ici considérait les actions comme de simples événements discrets, des unités indivisibles de comportement dont le contenu n'était pas défini.

2.2.1 La coordination du mouvement

Afin d'anticiper l'effort associé à un mouvement, nous devons donc décrire ce qui se passe précisément au cours d'un déplacement et trouver une mesure du coût de l'action qui reflète l'énergie investie pour bouger le corps de son point de départ vers sa destination.

Produire l'action même la plus simple nécessite de mobiliser un grand nombre d'articulations et de muscles afin de produire un déplacement coordonné des différentes parties du corps (Jeannerod, 1988). Maîtriser la mécanique de notre corps pour arriver à un objectif précis est ainsi une tâche difficile qui demande une parfaite synchronisation de l'appareil moteur tout au long de l'exécution du mouvement (Bernstein, 1967). Ainsi, nous ne pouvons plus nous limiter à décrire les actions comme des déplacements discrets car l'effort se déroule dans le temps, au rythme des activations musculaires qui permettent le déplacement.

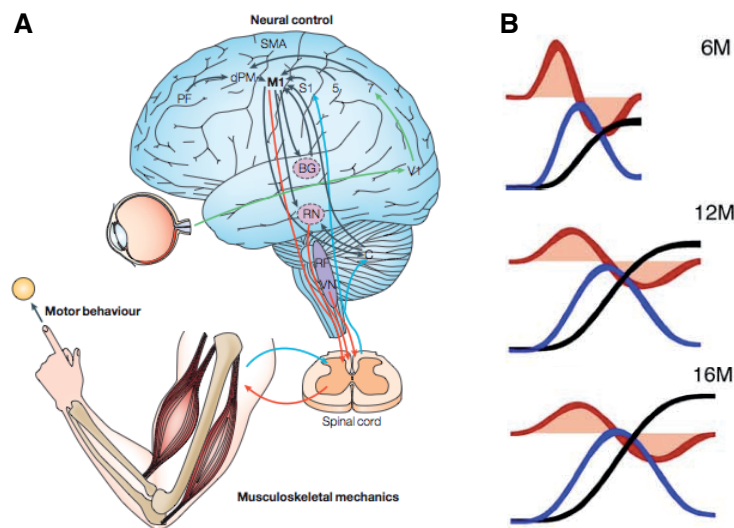


FIGURE 2.3 – *Le problème de la coordination.* **A** Le cerveau doit définir quelles commandes motrices envoyer aux muscles pour produire le comportement moteur le mieux adapté à la tâche (Scott, 2004). **B** Mesures empiriques (Mazzoni et al., 2007) de la cinématique du bras (noir : position, bleu : vitesse, rouge : accélération) lors de mouvements de pointage pour différentes amplitudes (6, 12 et 16 cm), révélant la nature des choix de contrôle par le cerveau.

Si le corps est une machine complexe, c'est le système nerveux central qui doit en maîtriser le pilotage (Wolpert, 1997; Wolpert et Ghahramani, 2000; Todorov, 2004). En effet, le cerveau doit tout d'abord élaborer des séquences précises de signaux nerveux (le *contrôle*) pour déclencher la contraction des muscles qui à son tour, par effet de levier, produira un mouvement du squelette (Kandel et al., 2000; Winter, 2009, voir figure 2.3). Ainsi, résoudre le problème de la coordination motrice, c'est déterminer le contrôle qui produira un comportement moteur en accord avec l'objectif visé. Le problème qui se pose ici est de trouver une méthode pour asservir la dynamique du corps à nos intentions, c'est-à-dire déterminer quelle séquence de signaux nerveux permettra d'arriver à la récompense et garantira le succès comportemental de l'action motrice.

2.2.2 Le principe du moindre effort

Une des difficultés que pose la coordination motrice est qu'il existe une multitude de façons différentes d'arriver à une position ciblée. Lorsque le singe de notre exemple a repéré un fruit au bout d'une branche, il peut sauter, marcher, utiliser ses mains ou ses pieds pour prendre appui, *etc.* : toutes ces possibilités lui permettront d'obtenir ce qu'il convoite.

S'ils partagent le même objectif, ces différents mouvements diffèrent cependant par le contrôle qu'ils requièrent. Certaines actions ne demanderont que peu d'activité musculaire, alors que d'autres nécessiteront de larges commandes motrices pour être réalisées. L'effort associé à l'exécution d'une action sera donc fortement lié au choix de mouvement, ou plus exactement au choix d'un contrôle particulier. Il est donc logique, à ce niveau, de proposer l'hypothèse que le cerveau cherche à travers ce choix à minimiser l'effort requis pour guider les mouvements.

Si l'hypothèse que l'action repose sur un principe de minimum d'effort fait intuitivement sens en terme de survie, elle permet également d'expliquer avec succès les

caractéristiques principales du contrôle moteur (Todorov, 2004; Guigon, 2011). L'idée d'utiliser un principe d'économie pour rendre compte de la coordination motrice a en effet été l'objet de nombreuses études au cours des dernières décennies.

Une question importante concernant ce modèle touche à la nature du coût optimisé (Engelbrecht, 2001) et donc à ce que nous entendons par « coût du mouvement ». De nombreuses fonctions de coût ont été proposées dans la littérature (Flash et Hogan, 1985; Uno et al., 1989; Harris et Wolpert, 1998; Todorov et Jordan, 2002; Guigon et al., 2007a). Ces approches sont toutefois toutes des variations autour d'un même principe de minimum d'effort, ce qui les rend d'ailleurs difficilement discernables sur la seule base de leurs prédictions qualitatives sur le comportement moteur (Engelbrecht, 2001; Guigon, 2011). Certaines de ces théories proposent que le coût de l'action peut être mesuré à partir de phénomènes périphériques, tel que l'énergie consommée par les muscles, alors que d'autres, comme la théorie du minimum de contrôle, postulent que le coût de l'action est évalué de façon centrale dans le cerveau. Cette deuxième hypothèse semble plus crédible (McCloskey, 1978), et nous nous alignerons ainsi avec les théories récentes du contrôle optimal en mesurant le coût de l'action à la taille des commandes motrices qu'elle requiert.

En définitive, nous avons proposé dans cette section l'hypothèse que le contrôle du mouvement suivait un principe de minimum d'effort, ou plus exactement de minimum de contrôle. La planification d'une action suivant ce principe permet d'évaluer l'effort que va coûter le mouvement en tenant compte de façon réaliste de la dynamique du corps.

2.3 Effort et durée de mouvement

Dans la première section, nous avons décrit le comportement comme une séquence d'actions discrètes qui permettent d'accumuler des récompenses au coup par coup. Nous avons mis en évidence dans la section suivante que cette description du comportement était insuffisante car elle ne prenait pas en compte le coût lié à l'exécution des actions. Afin d'y remédier, nous avons proposé un moyen explicite pour calculer le coût optimal d'un mouvement.

Il y a cependant une faille dans cette vision du comportement. Si nous avons insisté sur l'aspect continu du mouvement, nous avons toutefois défini l'action motrice comme un évènement limité dans le temps, c'est-à-dire comme un bout de comportement qui s'arrête quand la cible est atteinte. Ceci présente tout d'abord une sérieuse difficulté si nous voulons expliquer l'élaboration fluide et continue du comportement. En effet, les actions ne s'arrêtent jamais abruptement, et elles ne s'enchaînent pas non plus les unes après les autres sans pauses entre les phases de mouvement. En réalité, le découpage du comportement en actions distinctes est assez artificiel et ne reflète pas le fait que nous contrôlons notre corps de façon continue et permanente.

La durée apparente des mouvements est toutefois un élément clef pour comprendre le comportement. En effet, l'effort associé à une action dépend directement de sa vitesse d'exécution, car un mouvement plus rapide demande une plus grande force musculaire et aussi des commandes motrices plus importantes. Si nous tentons d'appliquer le principe de minimum d'effort sans contraindre la durée des mouvements, les actions seraient ainsi réalisées infiniment lentement.

Nous allons voir dans la suite de cette section que la durée des mouvements est un paramètre fondamental du comportement et qu'elle trahit les influences de l'effort et de la motivation sur l'élaboration du contrôle de l'action. Nous proposerons ensuite une solution pour faire émerger le temps de mouvement dans un modèle du comportement basé sur une régulation optimale de l'effort.

2.3.1 La régulation du temps de mouvement

Le premier facteur qui influence la durée des mouvements est la difficulté de l'action. Une manifestation de cet effet est l'augmentation de la durée des actes moteurs avec l'amplitude du déplacement. Cette caractéristique générale du comportement moteur, souvent appelée « loi de mise à l'échelle » se retrouve uniformément dans les mouvements du bras (Cooke, 1980; Brown et al., 1990; Gordon et al., 1994a; Hefter et al., 1996; van Beers et al., 2004) comme dans les saccades oculaires (Bahill et al., 1975; Van Opstal et Van Gisbergen, 1987; van Beers, 2007) ou même la locomotion (Hicheur et al., 2007; Bakker et al., 2007; Kunz et al., 2009; Mühlhoff et al., 2011). La sensibilité de la vitesse de mouvement à l'effort apparaît également au travers d'un effet de l'inertie du corps sur l'action (Gordon et al., 1994a; van Beers et al., 2004).

Différents facteurs externes vont également influencer la durée des mouvements. Il est par exemple possible de forcer des sujets à bouger plus vite simplement grâce à des instructions explicites (Cooke, 1980; Brown et al., 1990, cf. figure 2.4).

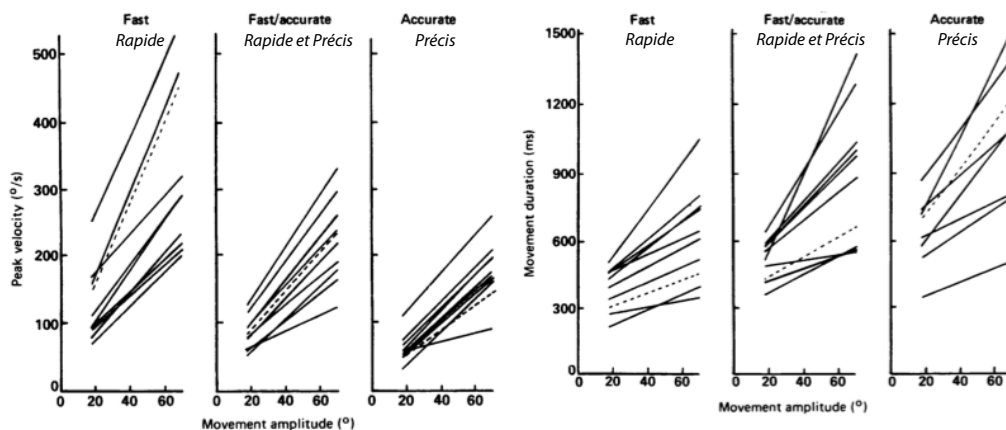


FIGURE 2.4 – *Modulation de la durée des mouvements en fonction de la tâche.* Vitesse (à gauche) et durée (à droite) de mouvements de flexion du bras pour différentes amplitudes (abscisse) et différentes instructions : rapide, rapide et précis, précis (Brown et al., 1990). Chaque ligne pleine représente un sujet, les lignes pointillées sont les moyennes du groupe.

Les humains sont également capables d'adapter la vitesse de leurs mouvements pour satisfaire des contraintes temporelles diverses, comme lorsqu'il leur faut attraper avec leur main une bille en train de chuter (Majsak et al., 1998; Ballanger et al., 2006, figure 2.5), ou bien identifier, suite à une saccade, une cible visuelle avant qu'elle ne disparaisse (Montagnini et Chelazzi, 2005).

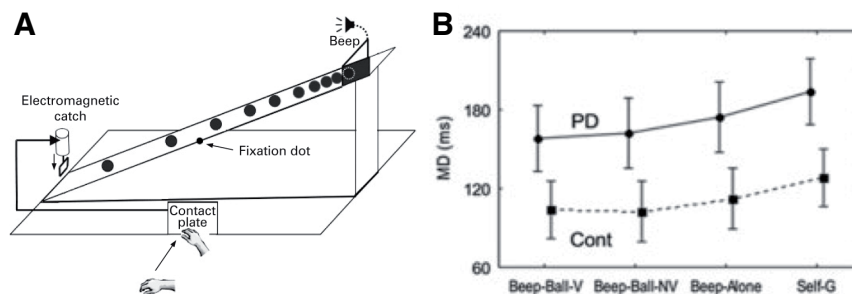


FIGURE 2.5 – *Situation d'urgence motrice* (Ballanger et al., 2006). **A** Des sujets ont pour instruction de toucher aussi vite qu'ils le peuvent un large bouton avec leur main, dans différentes conditions : quand ils se sentent prêts à agir (*self-generated*) ou en réaction à un signal auditif (*beep-alone*). Dans une dernière condition, une bille est lâchée sur une rampe inclinée au moment du signal sonore, et les sujets doivent toucher le bouton avant que celle-ci n'ait roulé en bas de la rampe (*beep-ball*). **B** Durée de mouvement pour les différentes conditions. Une différence significative entre les trois conditions se retrouve tant chez les sujets sains (*cont*) que chez les patients parkinsoniens (*PD*).

Un autre élément qui entre en jeu dans la régulation de la vitesse d'exécution de l'action est la motivation (figure 2.6). Par exemple, des rats qui ont faim courent plus vite vers de la nourriture que vers d'autres types de récompenses (Domjan, 2002; Peciña et al., 2003). Les singes font des saccades oculaires plus rapides vers les cibles qu'ils savent associées à une récompense (Takikawa et al., 2002; Watanabe et al., 2003), de même que les humains vers les cibles visuelles intrinsèquement plus intéressantes (Xu-Wilson et al., 2009). Plus généralement, la motivation, de même que la force des émotions (Schmidt et al., 2009), affecte la quantité d'effort investie dans une action (Pessiglione et al., 2007; Aarts et al., 2008).

Ce que révèle cette liste, loin d'être exhaustive, de phénomènes liés au temps

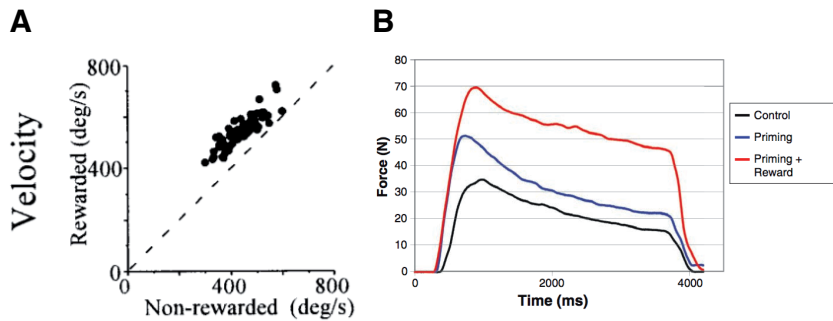


FIGURE 2.6 – *Effet de la motivation sur la production motrice.* **A** La vitesse des saccades oculaires vers des cibles récompensées (*rewarded*) sont plus rapides que vers des cibles non récompensées (*non rewarded*), chez le singe (Takikawa et al., 2002). **B** Des sujets humains qui ont subi une procédure de motivation implicite (*priming*) et éventuellement en plus une promesse de récompense (*priming + reward*) exercent spontanément plus de force dans une tâche de contraction isométrique de la main (Aarts et al., 2008).

des mouvements est que la vitesse de production des actions est une propriété du comportement moteur qui est régulée activement en fonction de la situation comportementale. Si nous voulons comprendre le comportement, il est donc indispensable d'expliquer comment nous choisissons la durée de nos actions et pourquoi nous la modulons en fonction de la tâche.

2.3.2 Le coût du temps

Essayons de voir en quoi le temps est un paramètre tellement important dans la production de l'action. Tout d'abord, une remarque importante est que la vitesse d'exécution des mouvements influence directement leur coût. En effet, pour aller plus vite, il faut utiliser plus de force et activer plus énergiquement ses muscles, ce qui implique que les mouvements rapides sont plus coûteux que les mouvements lents. Nous voyons apparaître là une limite de la théorie du moindre effort : si nous essayons d'appliquer à la lettre le principe de minimisation du coût du comportement, nous devrions nous déplacer le plus lentement possible.

La littérature du contrôle moteur élude souvent ce problème et se contente gé-

néralement d'imposer de façon arbitraire une durée maximale pour réaliser l'action. Quelques auteurs ont toutefois proposé des critères normatifs pour expliquer la façon dont peut être planifiée la durée des mouvements. L'idée habituellement avancée est que les actes moteurs minimisent, en plus du coût de l'effort, un « coût du temps » qui pénalise les mouvements trop lents (Hoff, 1994; Harris et Wolpert, 2006; Tanaka et al., 2006; Shadmehr et al., 2010a). La durée de l'action est alors planifiée pour optimiser le compromis entre ces deux coûts. Si cette approche présente de bonnes perspectives pour représenter le comportement moteur de façon descriptive, elle peine à s'intégrer dans un approche normative du comportement. Une première critique est qu'il est difficile d'expliquer conceptuellement quel est le prix du temps qui passe. En outre, cette définition de l'action ne permet pas de rendre compte explicitement de la modulation comportementale de la durée des actions, à moins d'expliquer comment la valeur du temps change en fonction de la tâche.

Une solution à ce problème nous vient de nombreux travaux d'économie et de psychologie qui proposent que si le temps n'a pas de coût dans l'absolu, il affecte la valeur anticipée des récompenses que nous visons (Frederick et al., 2002; Glimcher et al., 2007). Plus précisément, cette théorie postule que les récompenses se dévaluent avec l'attente qu'il est nécessaire d'endurer avant de les obtenir. Une interprétation possible de ce phénomène est qu'à trop tarder avant de se déplacer vers un objectif, nous risquons de perdre l'opportunité d'obtenir la récompense. Nous pouvons imaginer par exemple la situation où notre singe convoite un fruit à quelques branches de sa position actuelle. Comme d'autres singes sont, comme lui, à la recherche de nourriture, il doit se dépêcher s'il ne veut pas se voir dérober son fruit avant de l'avoir récupéré. Nous voyons dans cet exemple que la notion de dévaluation de la récompense capture l'idée que le temps ne coûte pas en lui-même mais implique un risque de perdre la récompense visée. Cette hypothèse permet d'expliquer pourquoi la durée des actions ne doit pas être trop longue, et plus précisément pourquoi la valeur des

récompenses peut influencer la durée de nos actions : plus l'objectif est important, plus grande sera la perte si nous prenons trop de temps pour agir.

Pour résumer, la sélection de la vitesse du mouvement doit faire face au dilemme suivant : d'un côté, le principe du minimum d'effort nous pousse à aller le plus lentement possible, alors que d'autre part le principe de maximisation des récompenses nous presse pour agir le plus vite possible pour limiter la dévaluation. S'il est impossible de satisfaire ces deux contraintes isolément, nous pouvons par contre faire l'hypothèse que ces deux influences sont en compétition lors de la planification du comportement. Un compromis émerge alors sous la forme d'une durée optimale d'action, une durée qui équilibre l'effort à investir dans l'action à la mesure de l'importance de l'objectif. Grâce à ce principe, nous pouvons maintenant comprendre comment notre singe prépare et planifie son comportement : il cherche à chaque instant à effectuer les actions les plus rentables, en étant prêt à investir plus d'énergie pour les récompenses les plus enviables pour ne pas risquer de les perdre tout en adaptant son effort à la difficulté de la situation pour ne pas s'épuiser plus que nécessaire.

2.4 Flexibilité du mouvement et du comportement

Dans un monde idéal, il nous suffirait d'appliquer à la lettre le comportement planifié pour obtenir ce que nous avons choisi. Toutefois, aussi précise soit la planification, de nombreux événements peuvent corrompre le déroulement de l'action. Revenons à notre singe au milieu de sa forêt pour illustrer les différents problèmes que le contrôle du comportement doit résoudre.

De façon générale, la façon dont réagit le corps n'est jamais exactement celle prévue, car la représentation de la dynamique corporelle qui sert à planifier l'action est par nature imparfaite. Tout d'abord, lorsque le cerveau de notre animal envoie des signaux de contrôle pour activer les muscles du corps, le déplacement qui en

résulte n'est pas toujours celui attendu. En effet, la transmission nerveuse est soumise à un bruit important qui dégrade la qualité du contrôle et entraîne une certaine variabilité dans l'exécution motrice (Messier et Kalaska, 1999; van Beers et al., 2004). Il peut également arriver à notre singe de surestimer sa force, à cause d'une fatigue périphérique soudaine par exemple. Plus fréquemment, les branches sur lesquelles il se déplace vont bouger de façon imprévisible et l'animal devra adapter son mouvement à ces forces qui le déstabilisent. Diverses études ont testé la capacité du contrôle moteur à réagir à ce type de perturbations mécaniques, en attachant un appareil robotisé au bras par exemple. Si ce genre de manipulations affecte grandement la coordination motrice, les sujets sont capables de corriger en ligne leurs mouvements afin d'atteindre leur but (Sanes, 1986; Shadmehr et Mussa-Ivaldi, 1994; Nijhof, 2003).

Notre singe doit également faire preuve de flexibilité pour s'adapter aux déplacements possibles de ses objectifs. S'il cherche par exemple à manger un insecte, il devra pour l'attraper être capable d'ajuster son mouvement en fonction des déplacements de sa proie. Là aussi, de nombreuses études ont montré que nous corrigeons nos mouvements de façon totalement fluide pour atteindre nos objectifs même s'ils se déplacent de façon imprévisible (Soechting et Lacquaniti, 1983; Goodale et al., 1986; Nijhof, 2003; Prablanc et al., 2003; Liu et Todorov, 2007).

Pour résumer, le processus de contrôle de l'action ne peut se contenter de préparer un mouvement *a priori* pour ensuite le laisser se dérouler sans intervenir : il doit réagir à chaque situation rencontrée pour produire dynamiquement les commandes motrices adaptées à la réalisation de l'objectif.

2.4.1 L'architecture du contrôle en ligne

Le principe selon lequel le contrôle du mouvement se fait en ligne est aujourd'hui largement accepté et bénéficie d'une assise théorique, neurophysiologique et computationnelle bien établie (Wolpert et Ghahramani, 2000; Scott, 2004, 2008;

Diedrichsen et al., 2010).

Ce principe repose sur une boucle fermée entre la partie du système nerveux en charge de la production des commandes motrices et l'appareil sensoriel qui constate l'effet de ces commandes sur le comportement moteur réel (figure 2.7) dont voici le schéma général : le *contrôleur* calcule un message de contrôle à envoyer aux muscles, le corps bouge selon sa dynamique propre, les différents capteurs sensoriels (proprioception, vision, *etc.*) évaluent le nouvel état du corps et de l'environnement, et en informent le contrôleur pour permettre l'ajustement du contrôle.

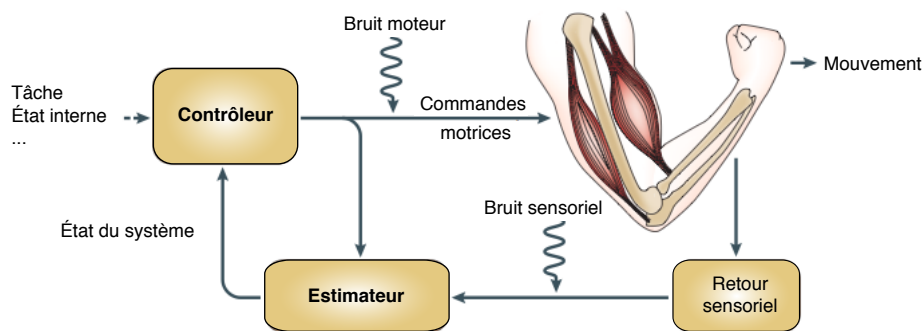


FIGURE 2.7 – Architecture générale du contrôle moteur. (Scott, 2004)

Tel quel, ce circuit action-perception a toutefois une grosse faiblesse. À cause des importants délais dans la transmission de l'information sensorielle, le contrôleur ne peut travailler que sur une représentation du corps périmée de plus de 100 *ms*, ce qui à l'échelle du comportement moteur peut faire une grosse différence (un mouvement de pointage ne dure que quelques centaines de millisecondes par exemple). Pour pallier les instabilités que peut générer ce délai, le cerveau semble utiliser, en parallèle de l'information perceptive, une simulation en interne du mouvement (Wolpert, 1997; Desmurget et Grafton, 2000; Shadmehr et al., 2010b). Si ce mécanisme de prédiction permet d'évaluer quasi-instantanément le résultat attendu des commandes motrices envoyées, il ne peut par définition rendre de compte des perturbations extérieures et imprévues qui affectent le mouvement.

Les informations sensorielles et les prédictions internes, complémentaires, sont finalement fusionnées au mieux par un processus d'*estimation* afin de former une représentation unique du corps sur la base de laquelle le contrôleur peut construire sa politique d'action (Körding et Wolpert, 2006).

2.4.2 Emergence de l'action

En définitive, le comportement n'est pas la simple application d'un plan d'action figé, il émerge d'une interaction permanente avec le monde qui nous entoure. Une conséquence directe de ce phénomène est que la durée des actions émerge également de l'ajustement continu des mouvements aux différents obstacles qui surgissent lors de l'exécution motrice. Sans remettre en cause le caractère planifié de la vitesse des mouvements, indispensable à une régulation optimale de l'effort, nous soulignons ici l'importance de la composante dynamique de la production de l'action. En particulier, toutes les expériences mettant en jeu des perturbations rapportent invariablement que la durée des actions est élastique, ce qui suggère que le temps de mouvement n'est pas un paramètre rigide du contrôleur mais plutôt une mesure relativement artificielle du comportement moteur apparent.

Cette remarque sape profondément les théories actuelles du temps de mouvement. Les méthodes habituellement proposées ne permettent en effet que de planifier une durée de mouvement finie, limitant à l'avance l'action sur un intervalle temporel rigide, et ne peuvent ainsi rendre compte de la grande flexibilité du contrôle du mouvement qui persévère dans l'action motrice tant que le but n'est pas atteint.

Le principe général, que nous avons proposé dans la section précédente, selon lequel le comportement repose sur l'optimisation du compromis entre effort restant et récompense à venir peut toutefois résoudre le paradoxe entre durée planifiée et durée réelle des actions.

Ce compromis peut en effet être ré-évalué à tout instant, et à chaque situation

rencontrée il fournit une marche à suivre, un contrôle à appliquer pour continuer à agir. Si aucune perturbation n'intervient, l'application répétée de ce principe est formellement équivalente à l'application aveugle du plan de départ. Si maintenant le moindre évènement déroge à ce qui était prévu, il est toujours possible de ré-évaluer le meilleur compromis pour la nouvelle situation et ainsi continuer à diriger le corps sans interruption. Plutôt que de découper le comportement en actions discrètes, nous proposons ainsi un principe pour diriger l'action motrice en continu et permettre ainsi l'accomplissement des objectifs visés en dépit des aléas rencontrés.

2.5 Un modèle général du comportement

Afin de vérifier que le critère du meilleur compromis entre effort et récompense permet effectivement d'expliquer les propriétés du comportement, nous allons construire formellement un modèle de l'action qui nous permettra de simuler la décision et le mouvement dans des situations réalistes. Dans cette section, nous nous appuyerons sur un exemple minimal pour bien comprendre les principes computationnels en jeu dans notre théorie.

2.5.1 Notion de dynamique

Nous adopterons l'approche des systèmes dynamiques pour modéliser le comportement (Jordan, 1996), c'est-à-dire que nous allons modéliser le corps comme un objet mécanique qui évolue sous l'influence d'un signal de contrôle optimal.

Commençons par décrire plus concrètement notre singe. Pour l'instant, nous pouvons le considérer comme un simple point défini par sa masse m et son *état* $\mathbf{x} = (p, \dot{p})$, *i.e.* sa position p et sa vitesse \dot{p} (le gras indique les vecteurs, le point la dérivée par rapport au temps). Ce singe se déplace sous l'effet d'une force F issue de l'activation

de son système musculaire par la commande motrice u envoyée par le cerveau :

$$F = h(u) \tag{2.1}$$

où la fonction h , dont nous ne précisons pas la forme pour le moment, représente la dynamique de la physiologie des muscles.

Nous pouvons maintenant décrire l'évolution de notre singe en expliquant comment cette force lui permet de se déplacer. Nous appliquons ici les lois de la mécanique classique qui nous indiquent que l'accélération \ddot{p} de l'animal sera égale à la force qu'il produit divisée par sa masse :

$$\ddot{p} = \frac{F}{m} \tag{2.2}$$

Ainsi, nous pouvons écrire plus généralement l'évolution de l'état \mathbf{x} du singe en fonction du contrôle u envoyé aux muscles :

$$\dot{\mathbf{x}} = \begin{cases} \dot{p} \\ \ddot{p} = h(u)/m \end{cases} \tag{2.3}$$

Formellement, nous parlerons de la *dynamique* du corps pour désigner les transformations physiologiques et mécaniques qui permettent de convertir un signal nerveux de contrôle \mathbf{u} (qui est de façon générale un vecteur) en un déplacement effectif $\dot{\mathbf{x}}$ du corps. Par souci de généralité, nous noterons simplement f cette dynamique :

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}) \tag{2.4}$$

Nous avons maintenant une description complète de notre animal. Son comportement apparent est décrit par la succession d'états qu'il traverse (sa position et sa vitesse au cours du temps). Ces changements d'états sont provoqués selon une dynamique précise (f) par l'application d'un signal de contrôle \mathbf{u} . La tâche de notre singe

(plus précisément de son cerveau) est maintenant de décider quel contrôle envoyer au corps pour produire le comportement le plus efficace.

2.5.2 Le principe d'optimalité

Nous allons maintenant formaliser l'idée que le choix du contrôle se fait dans l'optique d'obtenir le meilleur compromis entre effort et récompense. Cette écriture mathématique nous permettra de simuler notre modèle et ainsi de tester sa capacité à expliquer le comportement. Pour cela, nous allons adopter le cadre très général de la théorie du renforcement (Bertsekas, 1995; Sutton et Barto, 1998).

Bénéfice associé à un comportement

La première étape est de définir quelles sont les récompenses que l'animal espère trouver dans son environnement. Nous avons dit que le comportement de notre singe était décrit par les états \mathbf{x} qu'il visitait. Certains de ces états ne lui apportent rien (être assis sur une branche par exemple), alors que d'autres sont associés à une récompense particulière (attraper un fruit, *etc.*). De façon générale, nous noterons $R(\mathbf{x})$ la récompense que l'animal croit trouver à l'état \mathbf{x} donné. Si par exemple notre singe s'attend à obtenir une récompense de taille r s'il se déplace à une cible \mathbf{x}^* , nous pouvons écrire la fonction de récompense R ainsi :

$$R(\mathbf{x}) = \begin{cases} r & \text{si } \mathbf{x} = \mathbf{x}^* \\ 0 & \text{sinon} \end{cases} \quad (2.5)$$

Cette récompense peut être perçue comme étant plus ou moins intéressante selon le contexte ou l'état physiologique de l'animal. Nous utiliserons le paramètre ρ pour rendre compte de cet aspect subjectif de la valeur des récompenses. Ainsi, la récompense subjective associée à un état \mathbf{x} s'écrira $\rho R(\mathbf{x})$.

Nous pouvons maintenant écrire que le bénéfice global $J(\mathbf{u})$ d'un comportement

est égal à la somme des récompenses subjectives obtenues¹ en appliquant la politique de contrôle \mathbf{u} (Doya, 2000) :

$$J(\mathbf{u}) = \int_0^{\infty} e^{-t/\gamma} \rho R(\mathbf{x}_t) dt \quad (2.6)$$

où les différents états \mathbf{x}_t traversés au cours du temps résultent du déplacement occasionné, en fonction de la dynamique du corps, par la séquence de contrôle \mathbf{u} , *i.e.* en utilisant la dynamique $f(\mathbf{x}, \mathbf{u})$.

Le terme exponentiel est là pour représenter le mécanisme de dévaluation de la récompense au cours du temps. En effet, plus long est le temps t planifié pour arriver à l'état \mathbf{x}_t , plus petite sera la valeur prédite de la récompense (voir figure 2.8). L'importance de la dévaluation dépend du paramètre γ . Pour une durée donnée, plus petit sera ce paramètre, plus petite sera la valeur de la récompense dévaluée.

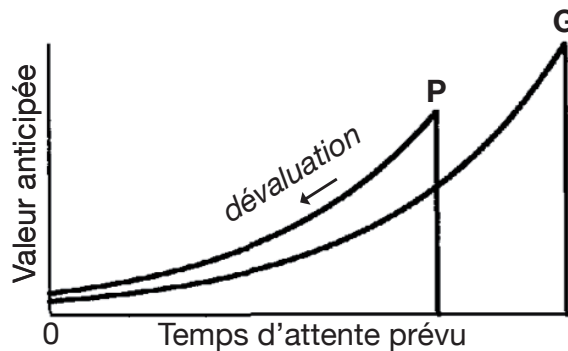


FIGURE 2.8 – *Illustration du phénomène de dévaluation.* Au moment présent, une grande récompense qu'il faut attendre longtemps (G) semble moins intéressante qu'une récompense plus petite mais plus rapide à obtenir (P).

Si à partir de là nous cherchons à optimiser le bénéfice de nos actions, il nous faut chercher la séquence de commandes motrices qui maximise le bénéfice total $J(\mathbf{u})$ du comportement. Cependant, comme jusqu'ici l'effort n'intervient pas dans le calcul de la valeur du comportement (équation 2.6), la solution à ce problème consiste à bouger

1. Afin d'être rigoureux mathématiquement, la fonction de récompense qui attribue la valeur r à un état \mathbf{x}^* doit être redéfinie, dans le cas continu, par $R(\mathbf{x}) = r\delta(\|\mathbf{x} - \mathbf{x}^*\|)$ où δ représente la fonction de Dirac.

le plus vite possible pour récupérer toutes les récompenses en un minimum de temps, sans considération pour les efforts que cela peut coûter. Nous allons alors intégrer le coût de l'action dans la planification de l'action.

Le coût de l'action

Nous avons proposé l'idée que le coût des mouvements était corrélé à la taille des commandes motrices qui servent à les produire. Dans notre formalisme, les signaux nerveux qui activent les muscles sont modélisés par le contrôle \mathbf{u} . Le coût instantané de l'action peut donc être évalué par la fonction de coût L définie par (Todorov, 2004; Guigon et al., 2007a) :

$$L(\mathbf{u}) = \|\mathbf{u}\|^2 \quad (2.7)$$

Comme pour les récompenses, nous modéliserons la valeur subjective de ce coût grâce à un paramètre de mise à l'échelle, η . Le coût subjectif associé à un contrôle \mathbf{u} est donc $\eta L(\mathbf{u})$.

Nous pouvons maintenant intégrer le coût moteur à notre modèle, en le soustrayant des bénéfices de l'action, pour obtenir une nouvelle définition de la valeur $J(\mathbf{u})$ d'un comportement :

$$J(\mathbf{u}) = \int_0^\infty e^{-t/\gamma} [\rho R(\mathbf{x}_t) - \eta L(\mathbf{u}_t)] dt \quad (2.8)$$

Cette équation nous dit que l'évaluation du comportement est possible grâce à une anticipation des récompenses et efforts subjectifs qui seront accumulés au cours du temps. Ainsi, optimiser le meilleur compromis entre les coûts et les bénéfices de nos actions, c'est trouver les commandes motrices que maximisent la valeur totale $J(\mathbf{u})$ du comportement, c'est-à-dire trouver le contrôle optimal \mathbf{u}^* qui résout :

$$J(\mathbf{u}^*) = \max_{\mathbf{u}} J(\mathbf{u}) \quad (2.9)$$

Si nous avons réussi à formaliser un principe général pour guider le comportement, il n'est pas évident à ce stade de comprendre comment maximiser concrètement le critère établi par l'équation 2.8.

2.5.3 Sélection d'une durée de mouvement optimale

Revenons à la situation qui nous pose problème. Une récompense de valeur subjective ρr se trouve à une position \mathbf{x}^* dans l'environnement. S'il veut maximiser ses bénéfices, notre singe doit se déplacer vers la cible \mathbf{x}^* . Pour cela, il doit trouver une séquence de commandes motrices \mathbf{u} pour produire à moindre coût les forces nécessaires au déplacement. La question qu'il reste à résoudre est de déterminer la quantité d'effort à investir dans cette action, c'est-à-dire définir à quelle vitesse exécuter le mouvement. Nous allons maintenant montrer comment le principe du compromis optimal (équations 2.8 et 2.9) permet de faire émerger une durée optimale de mouvement.

Admettons que l'animal planifie d'atteindre la cible à l'instant T . Pour y arriver à temps avec un minimum d'effort, il doit minimiser le coût $J_{\mathbf{u}}(T)$ du mouvement, c'est-à-dire la somme planifiée des commandes motrices à utiliser entre l'instant du départ et l'instant T où il arrivera en \mathbf{x}^* :

$$J_{\mathbf{u}}^*(T) = \min_{\mathbf{u}} J_{\mathbf{u}}(T) = \min_{\mathbf{u}} \int_0^T e^{-t/\gamma} L(\mathbf{u}_t) dt \quad (2.10)$$

Ce problème est très proche de la théorie du contrôle optimal classique (trouver le contrôle minimal pour un temps fini), et nous passerons pour l'instant sur la façon dont nous pouvons calculer effectivement ce coût (voir chapitre 3.1).

En planifiant de se déplacer en une durée T , le singe peut évaluer par la même occasion le bénéfice $J_r(T)$ de son action, en estimant la valeur de la récompense

dévaluée de la façon suivante :

$$J_r(T) = r e^{-T/\gamma} = \int_0^T e^{-t/\gamma} R(\mathbf{x}_t) dt \quad (2.11)$$

Au final, la valeur subjective totale $J(T)$ de l'action d'attraper la récompense en un temps T est égale à la différence entre le bénéfice attendu et l'effort anticipé.

$$J(T) = \rho J_r(T) - \eta J_u(T) \quad (2.12)$$

La valeur du comportement dépend ainsi de la durée de l'action. Plus le mouvement est lent, moins il est coûteux, mais plus la récompense est dévaluée. Inversement, un mouvement plus rapide limite l'effet du décompte mais demande un effort plus important. Il est toutefois possible de trouver une durée de mouvement optimale T^* pour laquelle le compromis entre ces deux influences est optimal et pour laquelle la valeur totale de l'action est maximale (figure 2.9) :

$$J^* = J(T^*) = \max_T [\rho J_r(T) - \eta J_u(T)] \quad (2.13)$$

Rappelons que les paramètres ρ et η pondèrent l'importance subjective accordée respectivement au gain et au coût de l'action. Autrement dit, ρ indique combien une récompense est récompensante et η combien un coût est pénalisant. Le paramètre γ , associé à la vitesse du décompte, détermine la façon dont la récompense est dévaluée avec le temps. L'ensemble des trois paramètres (ρ, η, γ) modélise l'appréciation subjective des valeurs en jeu (gain, coût, temps) et a un rôle capital dans l'évaluation du comportement et donc dans la sélection de l'effort à investir dans l'action. Ces paramètres seront appelés par la suite *paramètres internes* pour souligner leur nature subjective, par opposition aux paramètres de la tâche (la valeur et la position des récompenses par exemple) qui sont quant à eux externes à l'animal.

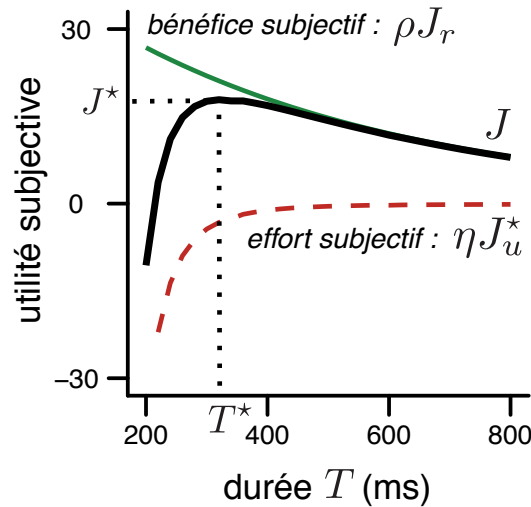


FIGURE 2.9 – *Emergence de la durée optimale de mouvement.* Le bénéfice (ligne continue verte) décroît avec la durée du mouvement. L’effort (ligne hachurée rouge), de signe opposé, diminue également. La valeur de l’action J (ligne épaisse noire) atteint son maximum à une durée T^* pour laquelle le compromis entre effort et bénéfice est optimal (pointillées).

Au final, le comportement optimal est de réaliser le mouvement en une durée T^* pour laquelle l’action atteint la valeur maximale J^* . Il faut noter que T^* pourrait ne pas exister dans le cas général. Toutefois, nous n’avons toutefois jamais rencontré ce cas dans la pratique. Le problème très général de trouver le meilleur compromis à long terme entre efforts et bénéfices (équations 2.8 et 2.9) peut ainsi se ramener à un problème de planification d’une action de moindre coût (équation 2.10) couplé à une recherche d’une durée optimale (équation 2.13).

Un point important de cette résolution est que nous obtenons deux informations sur l’action optimale. D’un côté, nous avons sa valeur subjective J^* qui synthétise les bénéfices et les coûts associés au mouvement. Cette utilité d’agir peut être calculée pour chaque récompense disponible dans l’environnement et servir de mesure subjective pour comparer les différentes options envisageables. D’un autre côté, nous obtenons pour un objectif donné un contrôle optimal pour exécuter l’action, c’est-à-dire une description explicite des commandes motrices à envoyer aux muscles pour

effectivement guider le corps vers le but visé.

2.5.4 Exécution en ligne du mouvement

Afin de simuler les tâches de décision, il suffit d'évaluer au moment du choix le bénéfice net anticipé pour chaque option disponible. Si nous voulons rendre compte de la coordination motrice dans des situations réalistes et étudier plus précisément le comportement moteur en présence de perturbations, nous devons par contre simuler la boucle d'action-perception qui caractérise le contrôle en ligne du mouvement. Dans cette section, nous décrirons brièvement ce mécanisme pour donner au lecteur une idée générale de son fonctionnement qui sera expliqué plus en détail dans la section 3.2.

Imaginons que nous venons de calculer le meilleur plan d'action étant donné la situation actuelle (position de départ, objectif, paramètres internes). Nous pouvons commencer à envoyer les commandes motrices au corps et simuler sa dynamique d'évolution en tenant compte des éventuelles perturbations mécaniques. Après un instant très court (1 ms virtuelle par exemple), nous simulons le processus d'estimation afin d'établir la nouvelle position du corps. Cette estimation peut alors servir de point de départ pour re-planifier l'action optimale à exécuter ensuite. Cette planification permet de redéfinir les commandes motrices optimales à envoyer au corps étant donnée la nouvelle situation. Ce cycle optimisation-exécution-estimation est ensuite répété, faisant émerger le comportement moteur. Le processus de contrôle en ligne peut ainsi être exécuté à l'infini. Cependant, une fois la cible atteinte, c'est-à-dire une fois le mouvement entièrement déroulé, le contrôleur ne fait que stabiliser le corps à sa position finale. La simulation peut alors être arrêtée.

Il est important de remarquer ici le rôle effectif du temps. Au départ nous avons critiqué les approches classiques du contrôle moteur qui reposent sur la spécification d'une durée *a priori* de l'action. Il peut sembler paradoxal que la démarche décrite ici repose essentiellement sur la détermination d'une durée de mouvement. Il faut com-

prendre que la durée en question n'est pas une durée prescriptive mais prospective, valide à un instant donné, sous l'hypothèse que les conditions courantes (*e.g.* récompense anticipée, valeurs subjectives des récompenses, des coûts et du temps) ne changent pas jusqu'à la fin du mouvement. De façon plus générale, la durée calculée n'est en fait qu'une quantité intermédiaire dans le processus d'optimisation. Un calcul direct des choix et des contrôles est possible sans passer par un calcul explicite de la durée. Par exemple, l'utilisation des techniques d'apprentissage par renforcement permet d'apprendre un tel calcul direct (Simpkins et Todorov, 2009; Marin et al., 2011).

Notre objectif dans la suite de ce document est de montrer que ce formalisme permet de rendre compte de la décision et du comportement moteur de façon réaliste. Le chapitre suivant décrit en détail la méthodologie employée pour simuler explicitement notre modèle.

Méthodologie

Ce chapitre relativement technique pourra être omis en première lecture, hormis peut-être la section 3.4 qui présente de façon détaillée les paramètres du modèle. Autrement, les éléments théoriques présentés dans le chapitre 2 devraient être suffisants pour comprendre les résultats.

3.1 Optimisation du contrôleur

Nous avons fait l'hypothèse que la maximisation de la valeur $J(\mathbf{u})$ sert de critère pour l'organisation des comportements moteurs, d'une part en attribuant à chaque action possible une utilité pouvant guider la décision, et d'autre part en prescrivant une politique de contrôle pour effectivement exécuter l'action. Afin de tester cette hypothèse dans des tâches réalistes, il est nécessaire de calculer explicitement ces valeurs. Ce chapitre décrit la méthodologie employée pour calculer l'utilité et le contrôle optimal pour un objet contrôlé générique de dynamique f (voir chapitre 3.3 pour plus de détails).

3.1.1 Résolution du contrôle optimal

Le premier problème est de trouver la séquence de contrôle \mathbf{u}^* capable de guider un système, défini par sa dynamique f , de son état initial \mathbf{x}_0 à l'état cible \mathbf{x}^* en une

durée T^* tout en minimisant la fonction de coût :

$$J_u = \int_0^{T^*} e^{-t/\gamma} L(\mathbf{u}_t) dt = \int_0^{T^*} e^{-t/\gamma} \|\mathbf{u}_t\|^2 dt \quad (3.1)$$

Ce problème est particulièrement difficile car le contrôle se développe dans le temps et que tout choix de contrôle réalisé à un instant donné a un impact sur tout le reste du mouvement. En effet, appliquer un contrôle \mathbf{u}_t va :

- augmenter immédiatement le coût du mouvement de $L(\mathbf{u}_t)$;
- faire bouger le système et éventuellement le rapprocher de la cible, diminuant alors le coût de la trajectoire restant à accomplir.

L'impact d'un contrôle instantané sur le coût global est donc ambigu. Pour pallier cette difficulté, nous allons nous doter d'un outil de mesure capable d'évaluer la qualité d'un contrôle à chaque étape du mouvement : l'Hamiltonien.

Formulation de l'Hamiltonien

Un contrôle \mathbf{u}_t apporte une contribution instantanée au coût du mouvement comme défini par le critère donné par l'équation 3.1, soit $e^{-t/\gamma} \|\mathbf{u}_t\|^2$.

Ce contrôle \mathbf{u}_t engendre également un déplacement $\partial\mathbf{x}_t$ du système selon la dynamique :

$$\frac{\partial\mathbf{x}}{\partial t} = f(\mathbf{x}, \mathbf{u}) \quad (3.2)$$

Soit $V(\mathbf{x}_t, \mathbf{u})$ le coût marginal de l'état \mathbf{x}_t , *i.e.* le coût de la trajectoire qu'il reste à parcourir à partir de l'état \mathbf{x}_t . Pour un déplacement élémentaire $\partial\mathbf{x}_t$ à partir de l'état \mathbf{x}_t , le coût marginal change d'une valeur $\partial V(\mathbf{x}_t, \mathbf{u})$. En notant $\boldsymbol{\lambda}_t = \partial V(\mathbf{x}_t, \mathbf{u})/\partial\mathbf{x}$, la contribution du contrôle \mathbf{u}_t au coût du trajet restant à parcourir est alors de $\boldsymbol{\lambda}_t^T f(\mathbf{x}_t, \mathbf{u}_t)$. La valeur de l'état adjoint $\boldsymbol{\lambda}_t$ peut être calculée à partir de l'état final

en utilisant (Pontryagin et al., 1986) :

$$\dot{\boldsymbol{\lambda}}^T = -\frac{\partial f}{\partial \mathbf{x}} \quad (3.3)$$

Nous pouvons maintenant écrire formellement l'Hamiltonien qui synthétise la contribution globale d'un contrôle au coût du mouvement :

$$\mathcal{H}(\mathbf{x}_0, \mathbf{u}, \boldsymbol{\lambda}, t) = e^{-t/\gamma} \|\mathbf{u}_t\|^2 + \boldsymbol{\lambda}_t^T f(\mathbf{x}_t, \mathbf{u}_t) \quad (3.4)$$

L'Hamiltonien résume ainsi la valeur économique à long terme d'un contrôle instantané \mathbf{u}_t (Dorfman, 1969 ; Sethi et Thompson, 1972, p34). Intuitivement, l'Hamiltonien est minimal pour le contrôle qui réduit le plus le coût marginal, et ce au moindre coût. Autrement dit, trouver le contrôle qui minimise le coût 3.1 en respectant la dynamique f , c'est donc trouver le contrôle qui minimise l'Hamiltonien. Cette propriété est appelée principe du minimum de Pontryagin (Pontryagin et al., 1986; Kirk, 2004), donné formellement par :

$$\frac{\partial \mathcal{H}(\mathbf{x}_0, \mathbf{u}, \boldsymbol{\lambda}, t)}{\partial \mathbf{u}} = 0 \quad (3.5)$$

Pour simplifier la suite de la résolution, débarrassons-nous du terme de décompte en utilisant le changement de variable $\boldsymbol{\xi} = e^{t/\gamma} \boldsymbol{\lambda}$, soit avec $\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{u}, \boldsymbol{\xi}) = e^{t/\gamma} \mathcal{H}(\mathbf{x}_0, \mathbf{u}, \boldsymbol{\lambda}, t)$:

$$\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{u}, \boldsymbol{\xi}) = \|\mathbf{u}\|^2 + \boldsymbol{\xi}^T f(\mathbf{x}, \mathbf{u}) \quad (3.6)$$

Notre problème de contrôle optimal revient ainsi à construire l'Hamiltonien et à en trouver le minimum, c'est-à-dire trouver le contrôle \mathbf{u} qui satisfait les conditions 3.2, 3.3 et 3.5, soit après le changement de variable :

$$\begin{cases} \frac{\partial \widetilde{\mathcal{H}}}{\partial \boldsymbol{\xi}} = \dot{\mathbf{x}} & (3.7) \end{cases}$$

$$\begin{cases} \frac{\partial \widetilde{\mathcal{H}}}{\partial \mathbf{x}} = -\dot{\boldsymbol{\xi}}^T + \frac{1}{\gamma} \boldsymbol{\xi}^T & (3.8) \end{cases}$$

$$\begin{cases} \frac{\partial \widetilde{\mathcal{H}}}{\partial \mathbf{u}} = 0 & (3.9) \end{cases}$$

Le cas de la dynamique linéaire

Commençons par essayer de résoudre le problème pour une créature simplifiée dont la dynamique peut se réécrire sous la forme :

$$f(\mathbf{x}, \mathbf{u}) = A\mathbf{x} + B\mathbf{u} \quad (3.10)$$

Dans ce cas, la dernière condition (eq. 3.5) permet de trouver explicitement le contrôle optimal :

$$\frac{\partial \widetilde{\mathcal{H}}}{\partial \mathbf{u}} = 0 = 2\mathbf{u} + \boldsymbol{\xi}^T B \implies \mathbf{u}^* = -\frac{1}{2} B^T \boldsymbol{\xi} \quad (3.11)$$

En remplaçant \mathbf{u} par \mathbf{u}^* dans la première condition (eq. 3.2), qui est en fait équivalente à la description de la dynamique (équation 3.10), nous obtenons la trajectoire optimale :

$$\dot{\mathbf{x}} = A\mathbf{x} - \frac{1}{2} B B^T \boldsymbol{\xi} \quad (3.12)$$

Cette dernière équation et la deuxième condition de Pontryagin (équation 3.3) forment alors le système d'équations différentielles :

$$\begin{pmatrix} \dot{\mathbf{x}} \\ \dot{\boldsymbol{\xi}} \end{pmatrix} = \begin{pmatrix} A & -\frac{1}{2} B B^T \\ -A & \frac{1}{\gamma} I_d \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\xi} \end{pmatrix} \quad (3.13)$$

dont il est facile d'exprimer la solution sous la forme :

$$\begin{pmatrix} \mathbf{x} \\ \boldsymbol{\xi} \end{pmatrix} (t) = \Gamma(t)C = \begin{pmatrix} \Gamma_{\mathbf{x}\mathbf{x}}(t) & \Gamma_{\mathbf{x}\boldsymbol{\xi}}(t) \\ \Gamma_{\boldsymbol{\xi}\mathbf{x}}(t) & \Gamma_{\boldsymbol{\xi}\boldsymbol{\xi}}(t) \end{pmatrix} \begin{pmatrix} C_{\mathbf{x}} \\ C_{\boldsymbol{\xi}} \end{pmatrix} \quad (3.14)$$

où C est un vecteur de constantes qui détermine la valeur du système aux bornes $t = 0$ et $t = T^*$.

Cette équation décrit la trajectoire optimale de notre créature au cours du temps. Par définition, cette trajectoire doit démarrer en \mathbf{x}_0 et se terminer en $\mathbf{x}_{T^*} = \mathbf{x}^*$. Plus formellement :

$$\begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}^* \end{pmatrix} = \begin{pmatrix} \Gamma_{\mathbf{x}\mathbf{x}}(0) & \Gamma_{\mathbf{x}\boldsymbol{\xi}}(0) \\ \Gamma_{\mathbf{x}\mathbf{x}}(T^*) & \Gamma_{\mathbf{x}\boldsymbol{\xi}}(T^*) \end{pmatrix} \begin{pmatrix} C_{\mathbf{x}} \\ C_{\boldsymbol{\xi}} \end{pmatrix} = \Gamma_{0T^*} \begin{pmatrix} C_{\mathbf{x}} \\ C_{\boldsymbol{\xi}} \end{pmatrix} \quad (3.15)$$

L'inversion de ce système donne analytiquement les constantes C . Finalement, la trajectoire optimale qui guide le système de sa position initiale à l'état cible sur l'intervalle $t = [0, T^*]$ est donnée par :

$$\begin{pmatrix} \mathbf{x} \\ \boldsymbol{\xi} \end{pmatrix} (t) = \Gamma(t)\Gamma_{0T^*}^{-1} \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}^* \end{pmatrix} \quad (3.16)$$

Le contrôle optimal \mathbf{u}^* peut alors être exprimé analytiquement en remplaçant le vecteur $\boldsymbol{\xi}$ ainsi calculé dans l'équation 3.11. Le coût total du mouvement J_u est ensuite calculable par intégration de l'équation 3.1.

Cas général

Si l'on modélise un système plus complexe, par exemple un bras articulé, la dynamique $f(\mathbf{x}, \mathbf{u})$ n'est plus linéaire. Les équations différentielles obtenues à partir des conditions de Pontryagin ne sont alors plus linéaires, ce qui rend leur résolution

analytique impossible dans le cas général.

Il faut donc recourir à des méthodes numériques pour déterminer la solution des conditions de Pontryagin. Divers algorithmes sont décrits dans la littérature du contrôle optimal (Bryson et Ho, 1975; Kirk, 2004). Plutôt qu'une description exhaustive, nous nous limiterons à un rapide survol de leur fonctionnement.

Pour une séquence de contrôle \mathbf{u} , la trajectoire de l'état \mathbf{x} et de l'état adjoint $\boldsymbol{\xi}$ peut être numériquement intégrée en utilisant les équations 3.2 et 3.3 à partir de \mathbf{x}_0 et de $\boldsymbol{\xi}_{T^*} = 0$. Nous pouvons alors construire le gradient de l'Hamiltonien le long de la trajectoire. Le minimum de l'Hamiltonien peut alors être approché en modifiant pas à pas le contrôle pour suivre ce gradient (méthode du gradient de premier ordre Bryson et Ho, 1975, p.221). Le contrôle est alors d'autant plus proche du contrôle optimal que l'on se rapproche du minimum de l'Hamiltonien.

3.1.2 Détermination du temps optimal

Nous venons de voir comment déterminer le contrôle optimal pour exécuter un mouvement entre \mathbf{x}_0 et \mathbf{x}^* en un temps quelconque T . Nous pouvons donc évaluer l'effort optimal $J_u^*(T)$ et ainsi la valeur $J(T)$ de l'action pour n'importe quelle durée de mouvement T . Il nous faut maintenant déterminer la durée T^* pour laquelle l'action atteint sa valeur optimale J^* .

Dans le cas où $J(T)$ peut être calculé analytiquement (dynamique linéaire), on peut simplement utiliser la condition :

$$\left. \frac{\partial J(T)}{\partial T} \right|_{T^*} = 0 \quad (3.17)$$

Dans le cas général, la seule méthode pour trouver la durée optimale est d'explorer la valeur que prend $J(T)$ pour différentes durées jusqu'à cerner approximativement la valeur de T^* , grâce à une méthode de bisection par exemple (Press et al., 1992).

3.2 Architecture générale du contrôle

Nous avons montré dans le chapitre précédent comment définir la politique de contrôle qui maximise le compromis entre efforts et récompenses. Utiliser cette politique pour effectivement diriger le corps dans l'environnement est toutefois une tâche délicate : l'action peut ne pas se dérouler comme prévu à cause du bruit inhérent au système ou bien de perturbations extérieures qui peuvent dévier le mouvement de son objectif. Pour garantir l'efficacité comportementale d'une politique, il est donc nécessaire de surveiller le déroulement des actes moteurs pour ajuster dynamiquement leur exécution. Nous allons voir ci-dessous comment contrôle et perception s'intègrent pour créer des comportements moteurs flexibles.

Un cadre normatif pour le contrôle moteur est l'architecture contrôle/estimation stochastique (Stengel, 1994; Todorov et Jordan, 2002; Todorov, 2004). La puissance explicative de ce cadre a été révélée par sa capacité à rendre compte de la formation des trajectoires pour des systèmes redondants (Todorov et Jordan, 2002; Guigon et al., 2007a), la correction en ligne des mouvements (Saunders et Knill, 2004; Liu et Todorov, 2007), la structure de la variabilité motrice (Todorov et Jordan, 2002; Guigon et al., 2008b), la formation des synergies (Todorov et Jordan, 2002), la loi de Fitts et le contrôle de la précision (Guigon et al., 2008b), les caractéristiques musculaires et neuronales des mouvements (Guigon et al., 2007b), et le contrôle simultané de la posture et du mouvement (Guigon, 2010). Il faut noter qu'une formulation stochastique du processus d'estimation est nécessaire, même si les simulations sont réalisées en l'absence de bruit (ce qui est le cas pour la plupart des résultats présentés ici). En effet, l'estimation (filtre de Kalman) est un processus qui pondère la simulation de l'effet des commandes motrices et des retours sensoriels sur la base de leurs variabilités respectives (Wolpert et al., 1995).

3.2.1 La boucle de contrôle

Distinguons trois modules à l'œuvre dans le processus de contrôle dans son ensemble, organisés comme montré sur la figure 3.1 ci-dessous :

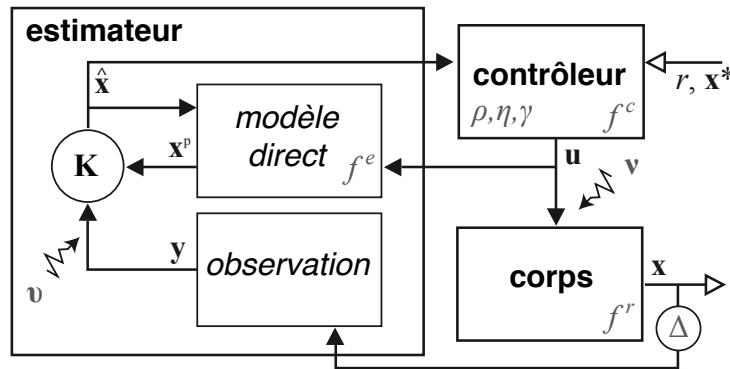


FIGURE 3.1 – *Modélisation de la boucle de contrôle.* Le contrôleur envoie les commandes motrices au corps qui produit alors un comportement moteur. L'estimateur se charge d'approximer la position du corps à partir des informations sensorielles et d'une copie des commandes motrices afin de permettre au contrôleur d'ajuster en ligne la politique de contrôle. Les variables des modèles sont indiquées en noir et les paramètres de chaque module en gris. \mathbf{K} désigne le filtre de Kalman, et Δ le délai sensoriel.

Le contrôleur a pour rôle de produire une commande motrice \mathbf{u} adaptée à la tâche (r, \mathbf{x}^*) étant donné l'état estimé du corps $\hat{\mathbf{x}}$. Ce processus est caractérisé par les paramètres internes (ρ, η, γ) et est optimisé pour une dynamique f^c (chapitre 3.1).

L'objet contrôlé est tout simplement le corps dont l'évolution est sous l'influence directe du contrôleur. Son état \mathbf{x} change suivant la dynamique réelle f^r .

L'estimateur quant à lui essaie de déterminer la position du corps pour en informer le contrôleur. L'estimation qu'il établit se base d'une part sur une observation passive \mathbf{y} grâce au système perceptif (proprioception, vision, *etc.*) et d'autre part *via* une anticipation active \mathbf{x}^p de l'effet du contrôle à l'aide d'un modèle interne f^e de la dynamique.

Ces trois processus agissent simultanément pour former une boucle de contrôle-déplacement-estimation d'où émerge le mouvement.

Bruit et perturbations

Le résultat de l'application d'une politique de contrôle est incertain. Une première raison à cela est que la dynamique pour laquelle est optimisé le contrôleur peut différer de la dynamique réelle du corps ($f^c \neq f^r$) ce qui provoque en général des mouvements inappropriés qui nécessitent des rectifications pour arriver à la récompense. Un autre problème inhérent à la physiologie du système nerveux est que du bruit corrompt inévitablement la commande motrice envoyée aux muscles. Ce bruit influence fortement le comportement moteur. Plus précisément, il apparaît que le bruit moteur augmente avec la taille du signal nerveux, principal phénomène à l'origine des propriétés statistiques de l'action motrice (Messier et Kalaska, 1999; van Beers et al., 2004). Pour rendre compte de façon réaliste de la variabilité du mouvement, nous avons modélisé le bruit moteur par un bruit multiplicatif $\nu \sim \mathcal{N}(0, \mathbf{u} \cdot \sigma_\nu)$ dont la variance est proportionnelle à la taille des commandes motrices :

$$\mathbb{E}[\nu\nu^T] = \sigma_\nu^2 \text{diag}(\mathbf{u}^2) \quad (3.18)$$

La présence de ce bruit multiplicatif est indispensable si nous voulons reproduire l'émergence de la variabilité caractéristique du contrôle moteur (Harris et Wolpert, 1998; Todorov et Jordan, 2002; Guigon et al., 2008b).

La perception est elle aussi soumise à du bruit. Ce bruit est surtout important pour déterminer l'influence relative de la perception dans le processus d'estimation, et n'affecte pas fondamentalement les propriétés statistiques du mouvement (Guigon et al., 2008b). La façon la plus simple de le modéliser tout en garantissant le bon fonctionnement de l'estimateur est sous la forme d'un bruit additif $\mathbf{v} = \mathcal{N}(0, \sigma_v)$.

La variance de ces bruits ne peut être mesurée expérimentalement, leur valeur

a donc été estimée pour que la variabilité des mouvements simulés corresponde aux observations empiriques. Les résultats exposés dans les chapitres suivants ont tous été obtenus avec les variances $\nu = 1$ et $\nu = .001$, la valeur exacte de ces paramètres n'ayant par ailleurs pas d'effet sur les propriétés du modèle.

3.2.2 Déroulement de la simulation d'une action motrice

Pour des raisons pratiques, il n'est pas possible de simuler la simultanéité d'exécution des divers processus de la boucle de contrôle. Il est cependant possible de l'approximer en simulant tour à tour le fonctionnement de chaque module sur un petit pas de temps Δ_t comme décrit ci-dessous. Par souci de clarté, nous n'explicitons toutefois pas la dépendance au temps des variables (état, contrôle, bruit) dans les équations ci-dessous.

Contrôleur

Imaginons que notre corps soit dans l'état \mathbf{x}_t , état dont nous pouvons nous faire une approximation $\hat{\mathbf{x}}_t$ grâce à l'estimateur. Le contrôleur planifie alors le contrôle optimal \mathbf{u}^* pour essayer d'atteindre un but (r, \mathbf{x}^*) visé à partir de cet état (voir chapitre 3.1).

Objet contrôlé

Le signal envoyé par le contrôleur est tout d'abord corrompu par le bruit moteur ν et la commande motrice réellement envoyée à l'objet contrôlé est :

$$\tilde{\mathbf{u}}^* = \mathbf{u}^* + \nu \quad (3.19)$$

L'évolution de l'état du corps (\mathbf{x}) est déterminée par la dynamique f^r , soit :

$$\dot{\mathbf{x}} = f^r(\mathbf{x}, \tilde{\mathbf{u}}) \quad (3.20)$$

C'est cet état effectif du corps qui correspond au comportement moteur réel et que nous mesurons objectivement dans les résultats.

Estimateur

L'estimateur prend alors le relais pour apprécier au mieux la position de l'objet contrôlé. S'il peut évidemment se reposer sur la perception pour évaluer l'état du corps, l'information fournie par l'observation est en général insuffisante pour garantir la coordination efficace du mouvement : la perception est un processus bruité et soumis à des délais de traitement longs en regard du déroulement de l'action (de l'ordre de 100 *ms*). De plus, l'appareil perceptif est composé de capteurs (les yeux, les fuseaux neuromusculaires, *etc.*) ne mesurant que partiellement l'état du corps (sa position et sa vitesse en particulier). Le cerveau semble pallier ces inconvénients en s'appuyant sur une simulation de l'évolution du corps grâce à une représentation interne (modèle direct, f^e) de la dynamique du système (Wolpert et al., 1995; Shadmehr et al., 2010b). Le mouvement imaginé est par contre aveugle aux perturbations qui, par définition, sont externes et imprévisibles. Afin de bénéficier des avantages de chaque type d'information, une solution est de les fusionner pour estimer le plus précisément possible l'état réel. Une méthode pour modéliser ce phénomène est d'implémenter un filtre de Kalman qui combine la perception et la prédiction interne en minimisant l'incertitude de l'estimation (van Beers et al., 2002; Todorov, 2005; Guigon et al., 2008a).

Le fonctionnement général de ce processus est de calculer l'incertitude respective de l'observation et de la prédiction interne et de pondérer ensuite chaque information par sa précision (Welch et Bishop, 1995). Cette pondération est calculée de façon optimale pour minimiser l'incertitude sur le résultat. Concrètement, l'estimation de l'état se déroule comme décrit ci-dessous.

Prédiction Grâce à un modèle interne f^e de la dynamique du corps, il est possible de prédire l'état du système étant donné le contrôle \mathbf{u}^* qui lui a été envoyé :

$$\dot{\mathbf{x}}^p = f^e(\mathbf{x}, \mathbf{u}) \quad (3.21)$$

A cause du bruit d'exécution et des perturbations, cette prédiction diffère de l'état réel du système d'une erreur \mathbf{e} donnée par l'équation :

$$\mathbf{e}^p = \mathbf{x} - \mathbf{x}^p \quad (3.22)$$

Observation Le système perceptif peut nous informer sur l'état réel et nous aider à corriger notre estimation pour réduire l'erreur \mathbf{e}^p . L'observation est donnée par :

$$\mathbf{y} = H\mathbf{x} + \mathbf{v} \quad (3.23)$$

où \mathbf{v} est un bruit gaussien de distribution $\mathcal{N}(0, R)$ et H est une matrice d'observation qui restreint l'observation à une sous-partie de l'état. En d'autres termes, la matrice H permet de ne garder que la partie observable de l'état et également de reproduire l'effet du délai perceptif (Todorov, 2005).

Ainsi, l'erreur de prédiction n'est mesurable que partiellement et imparfaitement. La partie de l'erreur qui peut être observée, *i.e.* l'erreur observable, est donnée par :

$$\tilde{\mathbf{e}}^p = H\mathbf{e}^p + \mathbf{v} = \mathbf{y} - H\mathbf{x}^p \quad (3.24)$$

Correction Si le système perceptif était totalement fiable, la meilleure solution serait de corriger notre prédiction de $\tilde{\mathbf{e}}^p$. Nous avons vu toutefois que l'appareil perceptif était très imparfait et que nous ne pouvons accorder un total crédit à l'information qu'il fournit. La correction de la prédiction par la perception doit donc être pondérée. Définissons à présent l'estimation *a posteriori* de l'état comme la prédiction ainsi

corrigée par l'observation :

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}^p + K\tilde{\mathbf{e}}^p \quad (3.25)$$

$$= \hat{\mathbf{x}}^p + K(\mathbf{y} - H\hat{\mathbf{x}}^p) \quad (3.26)$$

où K est une matrice gain (dans le cas scalaire, entre 0 et 1) qui représente la confiance relative accordée à l'observation. Si la perception est très bruitée, il vaut mieux ne pas en tenir compte et opter pour une correction qui sera très conservatrice ($K \rightarrow 0$). Si au contraire la perception est très précise, la prédiction aura intérêt à être pleinement corrigée pour correspondre à l'observation ($K \rightarrow 1$).

Après cette correction, l'erreur de prédiction *a priori* (entre l'état réel et la prédiction) sera réduite en une erreur d'estimation *a posteriori* (entre l'état réel et la prédiction corrigée) :

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} \quad (3.27)$$

Comme nous cherchons une estimation la plus précise possible, nous allons déterminer K pour qu'il minimise la variabilité de l'erreur d'estimation P :

$$P = \mathbb{E}[\mathbf{e}\mathbf{e}^T] \quad (3.28)$$

Pour trouver la valeur optimale de K , il suffit de substituer 3.26 dans 3.27 pour l'insérer ensuite dans l'équation précédente. En développant puis en annulant la dérivée par K , on obtient alors :

$$K = \frac{P^- H^T}{H P^- H^T + R} \quad (3.29)$$

où P^- est la variabilité *a priori* du processus de prédiction et R la variabilité de la

perception :

$$P^- = \mathbb{E}[\mathbf{e}^p \mathbf{e}^{pT}] \quad (3.30)$$

$$R = \mathbb{E}[\mathbf{y} \mathbf{y}^T] \quad (3.31)$$

La variance R est directement déterminée par les statistiques de l'observation et donc du bruit \mathbf{v} .

$$R = \mathbb{E}[\mathbf{y} \mathbf{y}^T] = \mathbb{E}[\mathbf{v} \mathbf{v}^T] \quad (3.32)$$

La variance de la prédiction *a priori* P^- peut quant à elle être calculée à partir de la variance *a posteriori* P . Il suffit pour cela d'utiliser les équations 3.20 et 3.21 pour déduire de l'équation 3.30 :

$$\dot{P}^- = f_{\mathbf{x}} P f_{\mathbf{x}}^T + f_{\mathbf{u}} Q f_{\mathbf{u}}^T \quad (3.33)$$

où $Q = \mathbb{E}[\mathbf{v} \mathbf{v}^T]$ est la variance du bruit d'exécution et

$$f_{\mathbf{x}} = \frac{\partial f^e}{\partial \mathbf{x}} \quad \text{et} \quad f_{\mathbf{u}} = \frac{\partial f^e}{\partial \mathbf{u}} \quad (3.34)$$

La prédiction corrigée $\hat{\mathbf{x}}$ est alors la meilleure estimation de l'état du corps. Cette estimation optimale peut en définitive être calculée à chaque instant à partir de l'estimation précédente et de l'observation courante :

$$\dot{\hat{\mathbf{x}}} = f^e(\hat{\mathbf{x}}, \mathbf{u}) + K(\mathbf{y} - H\hat{\mathbf{x}}) \quad (3.35)$$

La variance de cette nouvelle estimation est alors donnée par :

$$P = [1 - KH]P^- \quad (3.36)$$

Boucler la boucle

L'estimation optimale $\hat{\mathbf{x}}$ du nouvel état du système peut servir de départ pour la planification et l'application d'un nouveau contrôle, et ainsi de suite. Le contrôleur et l'estimateur forment ainsi un boucle de contrôle infinie en *feedback* (ou rétroaction) capable de s'ajuster en ligne aux diverses perturbations qui peuvent entraver le déplacement du corps et ainsi le bon déroulement du comportement.

Plus précisément, la simulation du modèle suit le principe suivant. A chaque instant t , l'état courant est \mathbf{x}_t et l'état estimé est $\hat{\mathbf{x}}_t$. Une politique de contrôle optimale est déterminée pour $\hat{\mathbf{x}}_t$ à partir de l'équation 2.13. Deux types de calcul sont nécessaires. Il faut tout d'abord savoir calculer $J_u(T)$, c'est-à-dire résoudre un problème de contrôle optimal à horizon fini. Ce problème se résout de façon analytique dans le cas linéaire (équation 3.15) et de façon numérique dans le cas non-linéaire (équations 3.7, 3.8 et 3.9). Il faut ensuite trouver la durée optimale du mouvement à partir de l'équation 2.13. Cela peut être déterminé de façon analytique dans le cas linéaire, ou à l'aide d'une méthode de recherche de minimum itérative, telle que la méthode de bisection (Press et al., 1992), dans le cas général. Les dynamiques de l'état réel (équation 3.20) et estimé (équation 3.35) sont ensuite intégrées (méthode d'Euler) entre les temps t et $t + \Delta_t$ pour la politique de contrôle \mathbf{u}_t^* choisie et les niveaux de bruit sensoriel et moteur $\sigma_v = 1$ et $\sigma_w = .001$. Nous obtenons alors le nouvel état réel ($\mathbf{x}_{t+\Delta_t}$) et estimé ($\hat{\mathbf{x}}_{t+\Delta_t}$) à partir desquels la procédure que nous venons de décrire peut être répétée. Le pas de temps d'intégration a été fixé à $\Delta_t = 1\text{ ms}$. La durée totale de la simulation est choisie empiriquement pour permettre un déroulement complet du mouvement.

3.3 Modélisation de l'objet contrôlé

Nous avons jusque là considéré un système extrêmement simple : une masse ponctuelle se déplaçant sur une ligne. Si cette simplification a suffi à l'exposition des principes qui nous intéressent, il est primordial de prendre en compte la richesse de la dynamique musculo-squelettique pour expliquer le comportement moteur dans des situations réalistes. En plus du modèle minimal utilisé jusqu'ici, ce chapitre donne deux descriptions formelles de la dynamique du bras à des niveaux croissants de détail. Cette gradation permettra lors des simulations de tester avec parcimonie les propriétés du modèle.

3.3.1 Objet I : point masse

Dans ce cas minimal, l'objet contrôlé est un simple point inertiel de masse m se déplaçant sur une ligne sous l'effet d'une force $h(u)$. La position p et la vitesse \dot{p} suffisent pour décrire l'état $\mathbf{x} = (\dot{p}, p)$ de ce système dont l'évolution est décrite par :

$$f_I(\mathbf{x}, u) = \dot{\mathbf{x}} = \begin{cases} \dot{p} \\ \ddot{p} = u/m \end{cases} \quad (3.37)$$

3.3.2 Objet II : point masse avec muscle

Lorsque le cerveau veut mouvoir le corps, il envoie par le biais de motoneurones des signaux aux muscles. Cette stimulation nerveuse déclenche la contraction des fibres musculaires visées et ainsi la production d'une force de traction qui fait bouger le squelette. Cette transformation d'une information électrique en énergie mécanique est appelée *couplage excitation-contraction* (Zajac, 1989; Kandel et al., 2000; Winter, 2009). Cette dynamique est importante car elle fait jouer aux muscles le rôle d'un filtre passe-bas qui lisse la commande motrice et contraint la production de force à une échelle temporelle caractéristique.

Une façon de capturer cette propriété est de représenter le muscle comme une fonction d'entrée-sortie agissant comme un filtre du deuxième ordre (Zajac, 1989; van der Helm et Rozendaal, 2000), comme schématisé sur la figure 3.2.A.

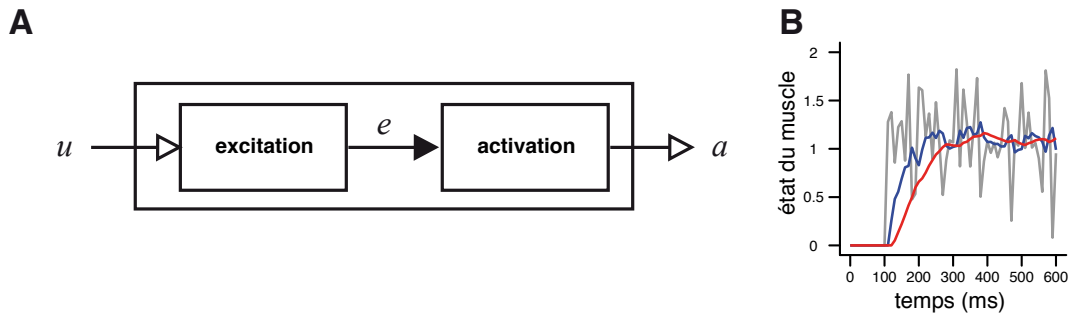


FIGURE 3.2 – *Modélisation du couplage excitation-contraction.* **A** Représentation du muscle comme un filtre. **B** Exemple d'un signal nerveux bruité u (gris) avec l'excitation e (bleu) et l'activation a (rouge) correspondantes pour $\omega = 1/.05$.

Formellement, le couplage excitation-contraction s'écrit

$$\begin{cases} \dot{e} = -\frac{e}{\omega_{e1}} + \frac{u}{\omega_{e2}} \\ \dot{a} = -\frac{a}{\omega_{a1}} + \frac{e}{\omega_{a2}} \end{cases} \quad (3.38)$$

où e et a sont respectivement l'excitation et l'activation musculaire, u est le signal nerveux envoyé par le cerveau *via* les motoneurones, et les différents ω , des constantes de temps. Par souci de simplicité, nous considérerons par la suite $\omega_{e1} = \omega_{e2} = \omega_{a1} = \omega_{a2} = \omega$. Une illustration du processus d'activation musculaire est donnée sur la figure 3.2.B.

L'objet I décrit dans la section précédente peut être facilement étendu pour prendre en compte la dynamique du couplage excitation-activation et nous fournir

un modèle minimal de la production de mouvement sous l'action d'un muscle :

$$f_{II}(\mathbf{x}, u) = \dot{\mathbf{x}} = \begin{cases} \dot{p} \\ \ddot{p} = a/m \\ \dot{a} = (e - a)/\omega \\ \dot{e} = (u - e)/\omega \end{cases} \quad (3.39)$$

L'objet II, tout en restant très simple à manipuler, permet de prendre en compte la latence et la propriété de filtrage des muscles. Il sera utilisé pour étudier l'effet du bruit sur la production motrice.

3.3.3 Objet III : bras articulé

La géométrie de notre corps influence également nos actions. Une répartition hétérogène des masses rend les mouvements plus ou moins aisés selon la position de nos membres et influe sur notre façon de bouger. L'implantation et les propriétés mécaniques des muscles participent également à la stabilité du mouvement face à des perturbations extérieures. L'objet III décrit ci-dessous est un modèle détaillé d'un bras humain et servira à tester nos hypothèses dans des situations réalistes de production motrice. Le réalisme physiologique a parfois été sacrifié afin de ne garder dans ce modèle que les propriétés bio-mécaniques ayant un rôle computationnel bien identifiable dans les simulations.

Géométrie du bras

L'objet III est un bras articulé à deux degrés de liberté (épaule et coude) dans le plan horizontal. Deux segments rigides représentent le bras et l'avant-bras, comme schématisé par la figure 3.3.

La position articulaire est donnée par $\boldsymbol{\theta} = [\theta_e, \theta_c]^T$. Chaque segment (numéro en indice) est décrit par sa longueur l , sa masse m , par la distance s de son centre de

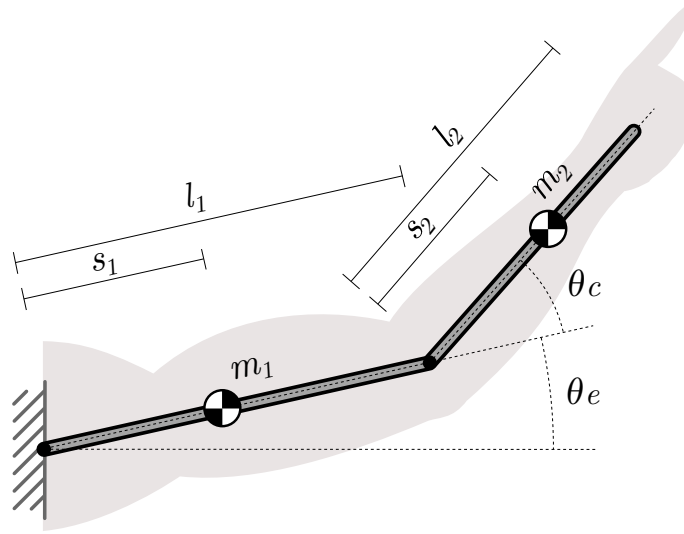


FIGURE 3.3 – Modélisation du bras par une chaîne articulée.

masse à l'articulation précédente ainsi que par son inertie I .

D'après la mécanique classique (seconde loi de Newton), nous pouvons décrire l'évolution de la position par :

$$\ddot{\boldsymbol{\theta}} = \mathcal{M}(\boldsymbol{\theta})^{-1} [\boldsymbol{\tau} - \mathcal{N}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}})\dot{\boldsymbol{\theta}}] \quad (3.40)$$

où $\mathcal{M}(\boldsymbol{\theta}) \in \mathbb{R}^{2 \times 2}$ désigne l'inertie du système et $\boldsymbol{\tau} = [\tau_e, \tau_c]^T$ le couple résultant de l'action des muscles. Les forces visqueuses $\mathcal{N}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}})$ peuvent être décomposées comme suit :

$$\mathcal{N}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) = \mathcal{C}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) + \mathcal{B} + \mathcal{J}(\boldsymbol{\theta})^T \mathcal{D} \mathcal{J}(\boldsymbol{\theta}) \quad (3.41)$$

où $\mathcal{C}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) \in \mathbb{R}^{2 \times 2}$ désigne les forces centrifuges et de Coriolis et $\mathcal{B} \in \mathbb{R}^{2 \times 2}$ les frottements articulaires. Un éventuel champ de force \mathcal{D} , exprimé dans l'espace de la tâche, se traduit dans l'espace articulaire grâce à la matrice jacobienne du système $\mathcal{J}(\boldsymbol{\theta})$.

Ces différentes matrices sont données par :

$$\mathcal{M}(\boldsymbol{\theta}) = \begin{bmatrix} \kappa_1 + \kappa_2 + 2\kappa_3 + \kappa_4 & \kappa_2 + \kappa_3 \\ \kappa_2 + \kappa_3 & \kappa_2 \end{bmatrix} \quad (3.42)$$

$$\kappa_1 = I_1 + m_1 s_1^2 \quad \kappa_3 = m_2 l_1 s_2 \cos(\theta_2)$$

$$\kappa_2 = I_2 + m_2 s_2^2 \quad \kappa_4 = m_2 l_1^2$$

$$\mathcal{C}(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) = \begin{bmatrix} \dot{\theta}_2 & \dot{\theta}_1 + \dot{\theta}_2 \\ -\dot{\theta}_1 & 0 \end{bmatrix} (-m_2 l_1 s_2 \sin(\theta_2)) \quad (3.43)$$

$$\mathcal{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \mathcal{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \quad (3.44)$$

$$\mathcal{J}(\boldsymbol{\theta}) = \begin{bmatrix} -l_1 \sin(\theta_1) - l_2 \sin(\theta_1 + \theta_2) & -l_2 \sin(\theta_1 + \theta_2) \\ l_1 \cos(\theta_1) + l_2 \cos(\theta_1 + \theta_2) & l_2 \cos(\theta_1 + \theta_2) \end{bmatrix} \quad (3.45)$$

Appareil musculaire

Si la conformation des masses du bras conditionne le mouvement et son contrôle, l'appareil musculaire participe également. Nous décrivons ci-dessous comment la configuration et la constitution des muscles participent au mouvement.

Configuration Lors de leur contraction, les muscles tirent sur les tendons par lesquels ils sont rattachés au squelette. C'est alors par un effet de levier que cette traction est transformée en force de rotation au niveau des articulations. Comme certains muscles ont leurs attaches sur des segments éloignés, ils agissent sur plusieurs articulations simultanément. Ceci crée des interactions inter-articulaires non négligeables que nous avons prises en compte dans notre modèle en considérant une paire de muscles bi-articulaires en plus des deux paires (un agoniste pour la flexion, un antagoniste pour l'extension) actionnant chaque articulation indépendamment (voir figure 3.4).

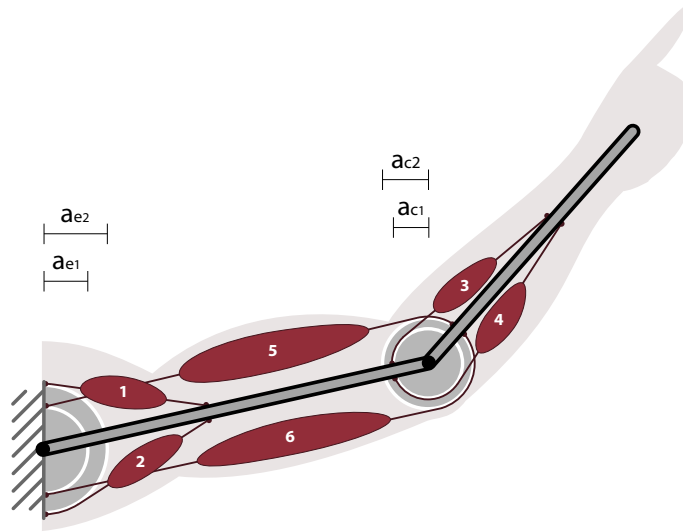


FIGURE 3.4 – *Implantation des muscles sur le squelette du bras.* Les muscles schématisés représentent respectivement 1) *Fléchisseur épaule* : deltoïde antérieur, pect. major, coracobrachialis 2) *Extenseur épaule* : deltoïde postérieur 3) *Fléchisseur coude* : biceps long, brachialis, brachioradialis 4) *Extenseur coude* : triceps latéral, anconeus 5) *Fléchisseur bi-articulaire* : biceps court 6) *Extenseur bi-articulaire* : triceps long

Selon cette architecture, le torque produit à chaque articulation peut alors être exprimé en fonction de l'ensemble \mathbf{F} des forces linéaires produites par chaque muscle :

$$\boldsymbol{\tau} = \mathcal{A}(\boldsymbol{\theta})^T \mathbf{F} \quad (3.46)$$

Le vecteur $\mathbf{F} \in \mathbb{R}^6$ contient les diverses forces musculaires et la matrice $\mathcal{A}(\boldsymbol{\theta}) \in \mathbb{R}^{2 \times 6}$ donne pour chaque muscle la distance de ses deux attaches par rapport aux articulations (c'est-à-dire leur moment articulaire). Par souci de simplicité nous considérons la matrice $\mathcal{A}(\boldsymbol{\theta})$ constante :

$$\mathcal{A}(\boldsymbol{\theta}) = \mathcal{A} = \begin{pmatrix} a_{e2} & -a_{e2} & 0 & 0 & a_{e1} & -a_{e1} \\ 0 & 0 & a_{c1} & -a_{c1} & a_{c2} & -a_{c2} \end{pmatrix}^T \quad (3.47)$$

Propriétés bio-mécaniques des muscles Tous les muscles du squelette possèdent des propriétés mécaniques similaires. Malgré la diversité de leur forme, ils

sont en effet tous constitués du même type de cellules, elles-mêmes contenant un grand nombre de compartiments contractiles identiques, les sarcomères. A l'intérieur, les filaments protéiniques qui permettent la contraction sont sensibles à la longueur et à la vitesse du muscle : ils produisent donc une force qui dépend de la cinématique du mouvement (cf. figure 3.5.A). Cette propriété bio-mécanique des sarcomères augmente la résistance visco-élastique des muscles et participe à la stabilisation du mouvement (van der Helm et Rozendaal, 2000).

En gardant la représentation du muscle comme une fonction entrée-sortie, la force F produite lors de l'activation du muscle peut se décomposer comme suit :

$$F = [a]^+ F_{max} F_L(\boldsymbol{\theta}) F_V(\dot{\boldsymbol{\theta}}) \quad (3.48)$$

où l'activation a est le résultat du couplage excitation-contraction décrit par l'équation 3.38, F_{max} désigne la force maximale que peut produire le muscle et les fonctions $F_L(\boldsymbol{\theta})$ et $F_V(\dot{\boldsymbol{\theta}})$ représentent respectivement les relations force/longueur et force/vitesse. La fonction $[a]^+ = \max(a, 0)$ garantit que la force reste positive, ce qui représente le fait que les muscles ne peuvent qu'exercer une traction lors de leur contraction.

Chaque muscle (en fait, chaque sarcomère) possède une longueur optimale à laquelle son potentiel de force est maximal. Si le muscle est trop court, les sarcomères sont tassés et ne peuvent plus se contracter. Inversement, si le muscle est trop allongé, les filaments contenus dans les sarcomères ne sont plus suffisamment en contact pour générer une traction. Ce phénomène est ici approximé par la fonction $F_L(\boldsymbol{\theta})$ montrée sur la figure 3.5.B, soit :

$$F_L(\boldsymbol{\theta}) = cL^2 - 2cL + c + 1 \quad (3.49)$$

où L est la longueur normalisée du muscle et c une constante. La longueur de l'en-

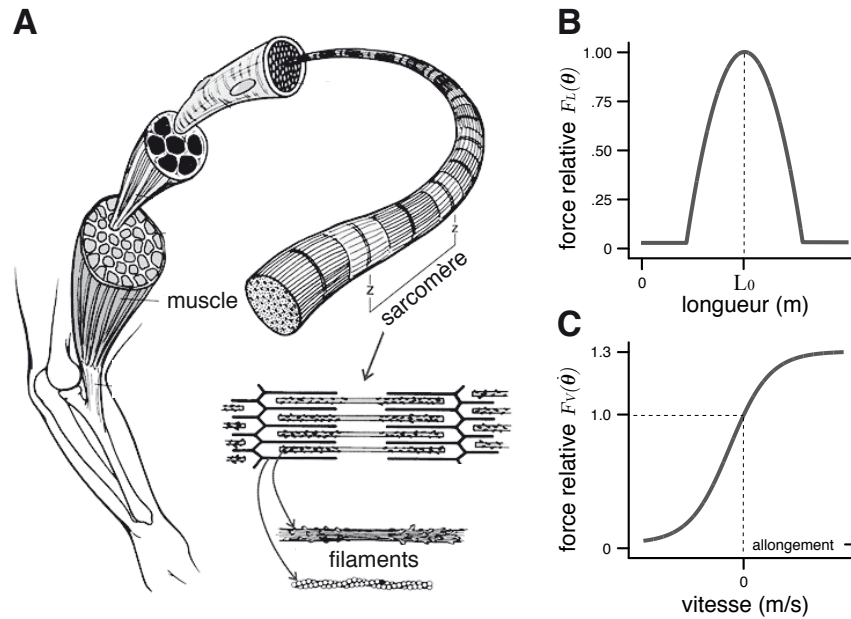


FIGURE 3.5 – *Constitution d'un muscle.* **A** Les muscles sont des faisceaux de fibres elles-mêmes constituées d'une multitude de compartiments, les sarcomères. À l'intérieur, les filaments protéiques interagissent pour permettre la contraction des sarcomères. La force produite par ces filaments dépend de leur configuration, plus exactement de leur longueur et de leur vitesse de déplacement. Ces propriétés se retrouvent à l'échelle du muscle. **B** Relation force/longueur. La force est maximale pour une longueur nominale L_0 . **C** Relation force/vitesse.

semble des muscles peut se calculer ainsi à partir de la position articulaire :

$$\mathbf{L} = (\mathbf{1} - \mathcal{A}\boldsymbol{\theta})/\mathbf{L}_0 \quad (3.50)$$

où \mathbf{L} est le vecteur des longueurs musculaires normalisées, \mathbf{L}_0 le vecteur des longueurs musculaires optimales pour chaque muscle et $\mathbf{1}$ le vecteur unitaire.

Lorsqu'un muscle est activé à une longueur inférieure à sa longueur optimale, la relation capturée par $F_L(\boldsymbol{\theta})$ est strictement croissante. Ainsi, pour un même niveau d'activation, un allongement mécanique du muscle entraîne une augmentation de la force et ainsi le raccourcissement du muscle. Cet effet de ressort permet aux muscles d'opposer une force élastique instantanée à toute perturbation extérieure qui viendrait perturber le mouvement.

La force contractile des sarcomères dépend aussi de leur vitesse d'allongement (ou de raccourcissement). Cette relation est approximée par la fonction :

$$F_V(\dot{\theta}) = \frac{d_1}{1 + d_2 \exp(-d_3 \dot{L})} \quad (3.51)$$

où d_1, d_2, d_3 sont des constantes et \dot{L} désigne la vitesse normalisée d'allongement du muscle que l'on peut calculer pour l'ensemble des muscles à partir de la vitesse articulaire :

$$\dot{L} = (\mathcal{A}\dot{\theta})/L_0 \quad (3.52)$$

Comme montré sur la figure 3.5.C, à un même niveau d'activation le muscle produira plus de force s'il est en cours d'allongement et moins de force s'il se raccourcit. Comme pour la relation force/longueur, ceci contribue à stabiliser le mouvement en opposant une force visqueuse instantanée à toute perturbation extérieure.

Comme nous l'avons vu, les muscles produisent de la force grâce aux fibres contractiles dont ils sont constitués. Plus il y a de fibres, plus gros est le muscle et plus grande est la force qu'il peut produire. Pour chaque muscle ou groupe de muscles, il est alors possible d'approximer la force maximale potentielle en multipliant leur grosseur (formellement, l'aire de leur section transversale *PCSA*) par une constante physiologique de tension *stress*. Autrement dit :

$$F_{max} = stress.PCSA \quad (3.53)$$

avec $stress = 350 \frac{N}{m^2}$ (Zajac, 1989). Les valeurs de *PCSA* et F_{max} sont données pour chaque muscle dans le tableau présenté en fin de chapitre.

Modèle complet

Pour résumer, l'état complet du bras est décrit par sa position et sa vitesse angulaire d'une part, et par le niveau d'excitation et d'activation de chacun des six muscles :

$$\mathbf{x} = [\dot{\theta}_e, \dot{\theta}_c, \theta_e, \theta_c, e_1, a_1, \dots, e_6, a_6]^T \quad (3.54)$$

La dynamique $\dot{\mathbf{x}} = f_{III}(\mathbf{x}, \mathbf{u}, t)$ est donnée par l'équation 3.40 pour la dynamique angulaire et, pour le couplage excitation-contraction de chaque muscle, par l'équation 3.38.

L'ensemble des paramètres pour chacun des muscles est résumé dans le tableau ci-dessous.

groupe musculaire	moments ep/co (m)	PCSA (cm^2)	F_{max} (N)	L_0 (m)
fléchisseur épaule	.040 / .000	22	700	.177
extenseur épaule	-.040 / .000	12	382	.229
fléchisseur coude	.000 / .025	18	572	.118
extenseur coude	.000 / -.025	14	445	.132
fléchisseur bi-articulaire	.028 / .035	5	159	.234
extenseur bi-articulaire	-.028 / -.035	10	318	.336

Pour les relations force/longueur nous utilisons $c = -1/.56$; pour la relation force/vitesse, nous utilisons $d_1 = 1.5$, $d_2 = .5$ et $d_3 = 5$.

3.4 Nature des paramètres

Avant de décrire les résultats, il est important de bien comprendre la nature des paramètres utilisés pour spécifier le modèle. Ces paramètres sont au nombre de cinq : \mathbf{x}^* , r , ρ , η et γ . Nous ne considérons pas ici les paramètres liés à la description des objets dont la nature est facile à comprendre.

3.4.1 Paramètres de la tâche

Le paramètre \mathbf{x}^* indique la localisation spatiale du but à atteindre et sert de condition aux limites pour l'élaboration de la politique de contrôle.

Le paramètre r est une valeur scalaire attachée au but qui peut correspondre à une récompense sur une échelle absolue (*e.g.* quantité de nourriture ou d'argent) ou à tout facteur qui module la poursuite et l'atteinte des buts (*e.g.* intérêt, probabilité, motivation, habitude). Ce dernier facteur reflète le fait que nous agissons dans un cadre bien plus large que celui défini par l'obtention de récompenses sonnantes et trébuchantes. Les paramètres \mathbf{x}^* et r sont des paramètres liés à la spécification de la tâche et seront appelés *paramètres de tâche*.

3.4.2 Paramètres internes

Pour décider et agir, une valeur de récompense doit être traduite dans une « monnaie » interne qui mesure combien une récompense est effectivement récompensante (paramètre ρ). Un sujet n'attribuera pas la même valeur à de la nourriture s'il est affamé ou repu, et la même valeur à l'argent s'il joue au Monopoly ou en bourse. Les paramètres r et ρ sont redondants, seul leur produit étant important (équation 2.8). Nous garderons cependant les deux paramètres car leur signification est différente.

Le paramètre η est un facteur de mise à l'échelle qui calibre la valeur subjective attribuée à un effort. Un sujet n'attribuera pas le même poids à un effort s'il est reposé ou fatigué. Les paramètres ρ et η sont redondants, seul leur ratio étant important (équation 2.8). Nous garderons cependant les deux paramètres car leur signification est différente. Ils peuvent aussi être l'objet de régulations différentes (*e.g.* niveau de richesse *vs.* niveau de fatigue). De façon générale, nous considérerons les variations du ratio ρ/η , que nous appellerons *vigueur* dans la suite.

Le paramètre γ est un facteur de dévaluation sur la récompense et l'effort. C'est à la fois un paramètre computationnel qui est nécessaire à la formulation du modèle, et

un facteur qui définit le processus par lequel les récompenses perdent de leur valeur au cours du temps (Green et Myerson, 1996; Stevens et al., 2005). Il faut noter qu'une diminution de γ induit une dévaluation plus grande.

Dans la suite, ρ , η et γ seront appelés *paramètres internes*, pour indiquer qu'ils ne sont pas spécifiés par l'environnement, mais servent de calibrage dans l'attribution de valeurs subjectives à des influences réelles liées au corps et à l'environnement. Ces paramètres seront autorisés à varier pour explorer leur rôle dans le modèle. Pour produire une interprétation neurale du modèle, nous proposerons des liens entre ces variations et les grandeurs physiologiques connues.

Le contrôle du mouvement

Nous voulons défendre l'hypothèse que le contrôle du mouvement et la décision émergent tous deux de l'optimisation du compromis entre l'effort et la récompense associés à une action. La première étape pour tester cette hypothèse est de vérifier qu'elle peut expliquer les propriétés du comportement moteur de façon réaliste. Les chapitres précédents indiquent comment calculer, à partir de notre hypothèse de départ, une politique de contrôle optimale capable de diriger la dynamique complexe du corps vers une récompense. Les résultats de simulation présentés ci-dessous montrent que le modèle permet non seulement de reproduire les caractéristiques fondamentales du mouvement (cinématique, statistiques) mais également d'expliquer la flexibilité comportementale de l'action motrice (modulation contextuelle de la durée, stabilité aux perturbations). Ces résultats démontrent ainsi que notre formalisme offre, grâce à des principes computationnels bien délimités, une description satisfaisante de la coordination du mouvement dans ses aspects mécaniques mais également comportementaux.

4.1 Propriétés fondamentales du mouvement

Nous avons montré dans le chapitre 2 que l'optimisation du compromis entre effort et récompense qui sous-tend notre théorie pouvait se ramener à un problème proche du principe d'effort minimal classique (équation 2.10). Cette équivalence suggère que les propriétés du mouvement qui émergent du principe du minimum d'effort devraient

également découler de notre modèle. Afin de le vérifier, nous avons tout d'abord testé la capacité du modèle à contrôler de façon réaliste un bras articulé en l'absence de bruit (moteur et sensoriel) vers des récompenses situées à différentes localisations et pour différentes valeurs de paramètres internes et externes.

4.1.1 Electro-myogramme

La mesure par électro-myogramme (EMG) de l'activité musculaire lors de la production de mouvements simples montre toujours une alternance tri-phasique de la paire de muscles agoniste/antagoniste de l'articulation concernée (Hallett et Khoshbin, 1980). L'EMG peut se mesurer dans notre modèle par la valeur absolue de l'excitation $[e]^+$ de chaque muscle. La figure 4.1 montre que le modèle reproduit cette caractéristique du contrôle moteur, un résultat qui se retrouve de façon robuste quels que soient les paramètres du contrôleur.

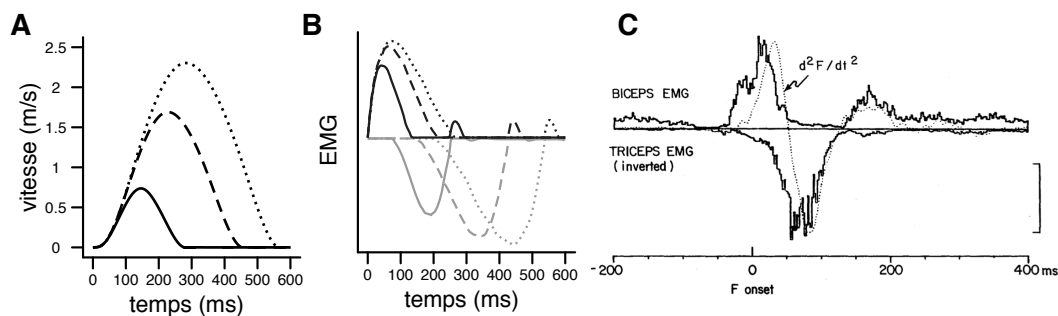


FIGURE 4.1 – *Electromyogrammes*. **A** Profils de vitesse simulés lors d'un mouvement de pointage vers une cible située respectivement à 10 (trait plein), 40 (tirets) et 70 cm (pointillés). **B** Activité électro-myographique correspondante de la paire de muscles agoniste (noir) / antagoniste (gris). Résultats obtenus avec le point masse avec muscles. Paramètres : $r = 1$, $\rho/\eta = 15$, $\gamma = 1$. **C** Exemple type du profil tri-phasique de l'EMG, enregistré ici pour une contraction isométrique chez l'homme (Gordon et Ghez, 1984).

4.1.2 Formation des trajectoires

Les trajectoires simulées sont montrées sur la figure 4.2. La main atteint la cible en suivant une trajectoire légèrement incurvée caractéristique des mouvements naturels. Le profil de vitesse de la main prend toujours la forme d'une cloche qui se met à l'échelle avec la durée du mouvement (figure 4.2.C) conformément aux résultats expérimentaux de la littérature (Gordon et al., 1994a).

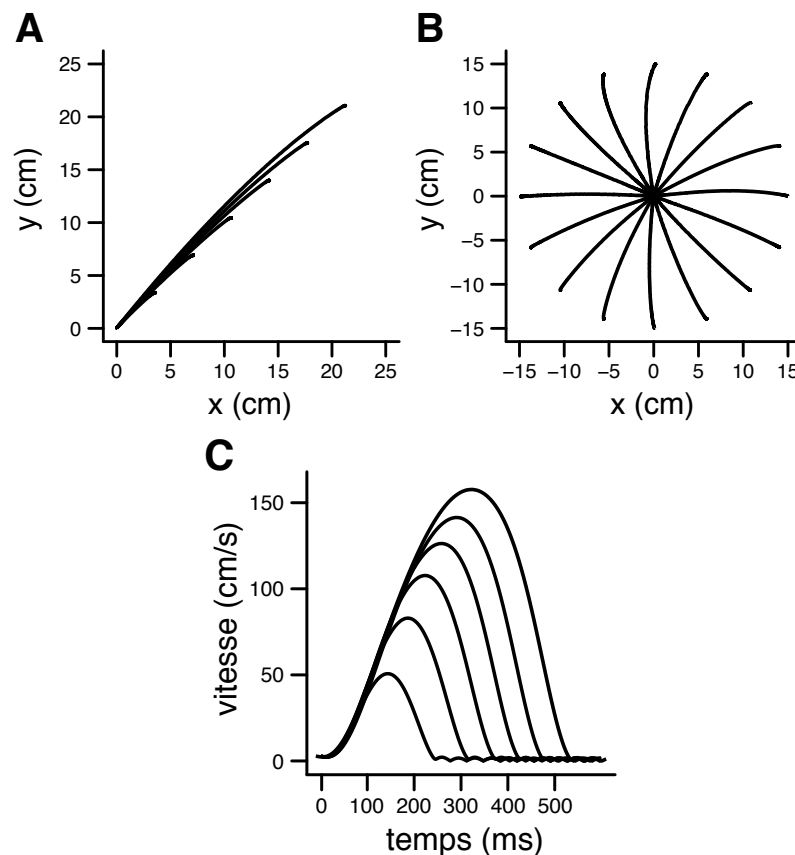


FIGURE 4.2 – *Trajectoires de la main.* **A** Cibles situées à 45° à différentes amplitudes (5, 10, 15, 20, 25 et 30 cm) **B** Cibles situées à 10 cm dans différentes directions. Position initiale du bras ($75^\circ, 75^\circ$). **C** Profils de vitesse correspondants aux mouvements montrés figure A. Résultats obtenus avec le bras articulé. Paramètres : $r = 40$, $\rho/\eta = 1/300$, $\gamma = .5$

Le principe d'effort minimal utilisé dans notre modèle permet ainsi d'expliquer quantitativement la cinématique et les EMG de mouvements réalistes, à l'instar de ce qui a été montré dans des travaux précédents (Todorov, 2004; Guigon et al., 2007a).

Le processus de contrôle que nous proposons se distingue toutefois des modèles classiques du contrôle optimal sur sa façon d'expliquer la durée des mouvements. Dans les formalismes existants, la durée de l'action est déterminée à l'avance de façon plus ou moins rigide (Todorov et Jordan, 2002; Harris et Wolpert, 2006; Tanaka et al., 2006; Shadmehr et al., 2010a). Ces modèles ne peuvent alors expliquer ce qui se passe après l'action : à la fin du temps imparti, le processus de contrôle « disparaît » tout simplement, que l'objectif soit atteint ou non. Dans notre cas, le contrôleur ne s'arrête jamais de diriger le corps et la durée de mouvement n'est finalement qu'une mesure extérieure du comportement moteur apparent. Après la phase de déplacement, qui dure tant que la cible n'est pas atteinte, la politique de contrôle continue en effet de diriger le corps et le stabilise jusqu'à ce qu'un nouvel objectif se présente (voir figure 4.2.C). Notre modèle résout ainsi naturellement la question du passage du mouvement à la posture en unifiant ces deux processus sous la forme d'un unique contrôleur en ligne. La théorie du contrôle postural prédictif proposée par Guigon (2010) est d'ailleurs basée sur un modèle très proche du formalisme présenté ici. Toutefois, une étude plus approfondie de notre modèle est nécessaire pour confirmer son aptitude à expliquer le contrôle postural.

4.2 Effort et temps de mouvement

Une particularité de notre formalisme est de planifier la durée du mouvement en fonction d'un compromis entre la récompense décomptée et l'effort requis pour atteindre la cible (équation 2.13). Notre modèle prédit ainsi que le temps de mouvement doit être modulé par la difficulté associée à la réalisation de la tâche. En effet, le processus d'optimisation aura tendance à planifier une durée du mouvement plus longue pour compenser un déplacement autrement trop coûteux (à cause la masse du corps, ou de la distance à parcourir par exemple). Cette propriété donne un sens aux

fluctuations de la durée de l'action liées à la direction et à la longueur du mouvement.

4.2.1 Loi de mise à l'échelle

La figure 4.2.C montre que la forme du profil de vitesse d'un mouvement ne varie pas en fonction de la position de la cible. Il apparaît toutefois que la durée des mouvements simulés augmente avec la distance à la récompense. Ce phénomène est observé expérimentalement de façon robuste et semble trahir un fonctionnement fondamental du contrôle moteur. Il est usuellement décrit par une « loi de mise à l'échelle » qui relie linéairement la durée des mouvements à leur longueur. Pour vérifier que cette loi émerge de notre modèle, nous avons mesuré la durée de mouvements simulés pour des cibles situées à différentes distances. Afin de nous conformer à la littérature expérimentale, la fin de chaque action est déterminée comme le premier instant où la vitesse passe en dessous d'un seuil arbitraire de 3 cm/s . La figure 4.3 montre que la loi de mise à l'échelle émerge correctement de notre formalisme. Ce résultat n'est pas affecté qualitativement par les autres paramètres du modèle (mais voir chapitre 6).

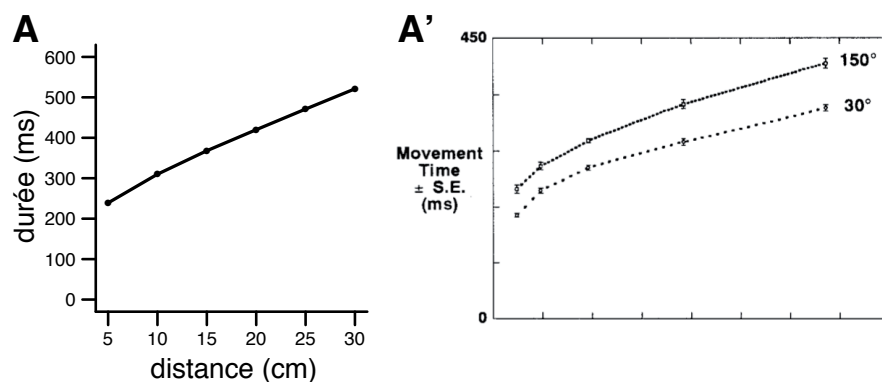


FIGURE 4.3 – *Loi de mise à l'échelle*. **A** Durée des mouvements en fonction de leur amplitude respective. Mêmes paramètres que sur la figure 4.2.A et C. **A'** Exemple de lois de mise à l'échelle observées expérimentalement (Gordon et al., 1994a).

4.2.2 Anisotropies inertielles

Certains mouvement sont plus à faciles à réaliser que d'autres car la mobilité de notre corps est influencée par la configuration articulaire et l'action des muscles (points d'insertions, bras de levier). Bouger notre avant-bras est par exemple moins gourmand en énergie que bouger notre bras tout entier, les masses en jeu n'étant évidemment pas les mêmes. A distances équivalentes, des cibles demanderont ainsi des efforts différents. Cette différence d'inertie apparente en fonction de la direction de déplacement (l'anisotropie inertielle) affecte la durée des actions : plus la masse à déplacer est grande, plus lent est le mouvement (Gordon et al., 1994a). Ce phénomène émerge également de notre modèle (figure 4.4). Ceci s'explique comme pour la loi de mise à l'échelle par le fait que la difficulté objective du mouvement affecte le compromis temporel trouvé par le processus d'optimisation entre effort et récompense de notre modèle.

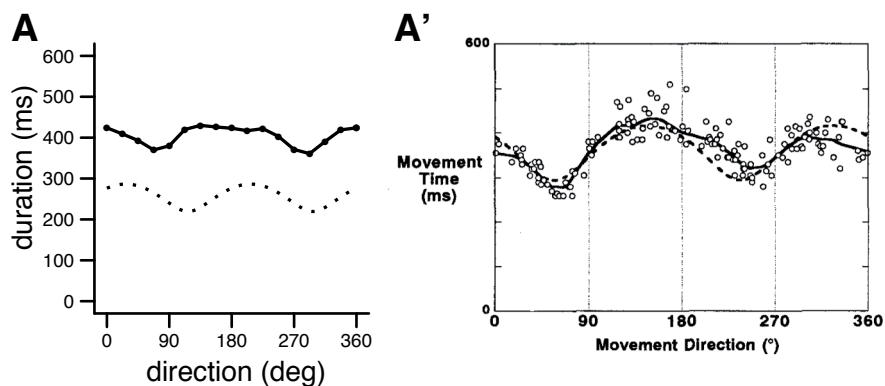


FIGURE 4.4 – *Effet de l'anisotropie inertielle*. **A** Durée des mouvements en fonction de leur direction. Mêmes paramètres que sur la figure 4.2.B. La ligne pointillée représente l'inertie apparente du bras en unité arbitraire. **A'** Données expérimentales de Gordon et al. (1994a).

Notre formalisme propose en définitive une explication unifiée de l'impact de la difficulté d'une action sur sa durée d'exécution, que cette difficulté soit imputable aux capacités de notre système moteur (*e.g.* son anisotropie inertielle) ou à la tâche

à surmonter (*e.g.* la distance à parcourir). Les deux résultats présentés ici sont des exemples de la capacité plus générale de notre modèle à moduler de façon prédictive la durée des mouvements en fonction du travail qui s'oppose à l'obtention d'une récompense (Gordon et al., 1994a), durée de mouvement qui émerge du meilleur compromis entre effort et récompense (équation 2.13). Nous verrons dans le chapitre 6 comment les aspects motivationnels (les paramètres internes) influencent également la régulation du temps de mouvement.

4.3 Statistiques du mouvement

Le chapitre 3.2 explique comment une méthode de planification optimale, une fois couplée à un processus d'estimation, peut former un processus de rétro-contrôle capable de diriger un mouvement vers son but malgré les déviations qui peuvent altérer l'exécution motrice. En l'absence de perturbations extérieures, le contrôleur doit déjà gérer le bruit interne qui dégrade la commande motrice. Ce bruit, qui augmente avec la taille du contrôle, est la principale source de la variabilité des actes moteurs (Harris et Wolpert, 1998; van Beers et al., 2004).

Divers travaux ont montré que la théorie du contrôle optimal arrivait à expliquer la structure de la variabilité motrice (Todorov et Jordan, 2002; Guigon et al., 2008b,a; Valero-Cuevas et al., 2009). En effet, en minimisant la quantité de contrôle nécessaire à la conduite du mouvement, un contrôleur optimal minimise indirectement le bruit moteur (voir équation 3.18). Notre modèle s'appuie également sur ce principe, ce qui explique sa faculté à reproduire également la variabilité des mouvements avec réalisme.

Afin de tester concrètement le formalisme proposé, nous avons simulé un grand nombre de répétitions ($n=300$) de différents mouvements en présence de bruit (moteur et sensoriel, cf. chapitre 3.2).

4.3.1 Profil de variabilité

Tout d'abord, nous avons mesuré la variabilité positionnelle du système au cours du temps comme montré sur la figure 4.5 pour un mouvement de pointage simple. Le modèle produit le profil caractéristique de variabilité des actions motrices réelles (Osu et al., 2004; Selen et al., 2006; Morishige et al., 2006). Plus particulièrement, ce résultat montre que notre formalisme suit le principe de minimum d'intervention qui caractérise le contrôle moteur : la variabilité au milieu du mouvement est plus importante qu'à la fin de l'action (ici vers 600 *ms*). Si nous laissons le contrôleur diriger le mouvement bien après la fin de l'action, la variabilité positionnelle revient progressivement à son niveau de départ, après un léger rebond au moment de l'arrêt apparent du mouvement. Les données présentées dans la littérature sont toutefois toujours coupées très tôt à la fin du mouvement (souvent car l'action est interrompue par un contact de la main avec la cible) et il est difficile d'évaluer objectivement le comportement du modèle dans la partie finale de l'action.

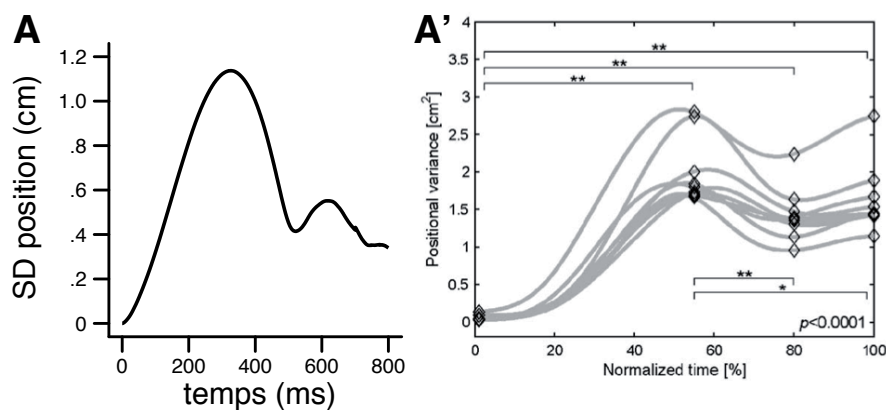


FIGURE 4.5 – *Profil de variabilité*. **A** Ecart type de la position de la main au cours du temps lors d'un mouvement de pointage. Simulations obtenues avec le point masse avec muscles pour une distance de 30 *cm* avec $r = 1$, $\rho/\eta = 100$ et $\gamma = 2$. **A'** Données expérimentales de Morishige et al. (2006) pour une tâche similaire.

Les mouvements d'amplitude croissante sont exécutés à des vitesses grandissantes (loi de mise à l'échelle) et demandent donc une quantité de contrôle plus grande.

Ainsi, les mouvements plus grands sont soumis à un bruit moteur plus important. La conséquence de ce phénomène est que les mouvements de plus grande amplitude sont plus variables spatialement (Searle et Taylor, 1948; Prablanc et al., 1986; Worringham, 1991; Gordon et al., 1994b; Khan et al., 2004; van Beers et al., 2004; van Beers, 2007). Contrairement aux approches existantes, notre modèle ne fixe pas à l'avance la durée du mouvement mais la fait émerger du processus de contrôle en ligne de l'action. Les simulations montrent ainsi que le bruit moteur peut aussi expliquer la variabilité de la durée des mouvements. En particulier, notre modèle reproduit l'augmentation de la variabilité temporelle des actions avec l'amplitude des mouvements (Searle et Taylor, 1948; Brown et Slater-Hammel, 1949; Nagasaki, 1989; van Beers, 2007), comme un corollaire direct de la loi de mise à l'échelle.

Les propriétés précédentes émergent du modèle quelle que soit la valeur du bruit (dans des limites raisonnables). De façon similaire à des observations précédentes (Guigon et al., 2008b), c'est le rapport entre bruit moteur et bruit sensoriel (ν/ν) qui semble déterminer principalement l'amplitude absolue de la variabilité au cours et la fin du mouvement. Si ce phénomène n'a toutefois pas fait l'objet de tests exhaustifs, il est intéressant de noter que le bruit sensoriel paraît être un déterminant important de la variabilité des mouvements (Guigon et al., 2008b).

4.3.2 Loi de Fitts

La précision des mouvements dépend fortement de la vitesse à laquelle ils sont réalisés. Les actions les plus rapides sont en effet les moins précises, principalement à cause du bruit moteur qui augmente avec la taille des commandes motrices envoyées aux muscles. Il existe ainsi une relation forte entre l'amplitude, la durée et la précision d'une action, un phénomène robuste capturé par ce qui est communément appelé « loi de Fitts » en référence à son découvreur (Fitts, 1954). Cette loi indique que la durée d'un mouvement augmente linéairement avec son « indice de difficulté » (ID), ce

dernier étant calculé comme le rapport entre l'amplitude du mouvement (A) et la taille de la cible (W) : plus une action est grande et/ou précise, plus elle est lente.

Notre modèle ne permet pas de spécifier la vitesse d'un mouvement directement à partir d'une taille de cible donnée. Nous ne pouvons donc simuler directement les expériences classiques de la loi de Fitts où la précision est imposée par la tâche et où la durée de mouvement est mesurée empiriquement. Pour vérifier si notre modèle reproduit la loi de Fitts, nous allons renverser le problème en modulant la vitesse de mouvement et en mesurant W à la variabilité effective de l'action (Bainbridge et Sanders, 1972). Afin de faire varier la vitesse d'exécution d'une action dans nos simulations, il nous faut jouer sur les différents paramètres internes du modèle ρ , η et γ (cf. chapitre 6 pour plus de détails). En déterminant, comme précédemment, la fin de chaque mouvement au premier instant où la vitesse descend en dessous de 3 cm/s , nous pouvons mesurer la variabilité W du point d'arrivée (en fait l'intervalle de 95% de confiance) pour chaque amplitude (A) de mouvement et pour chaque ensemble de paramètres. La figure 4.6.A montre que la durée de mouvement et l'indice de difficulté du mouvement ($ID = \log(A/W)$) sont bien liés de façon linéaire et que le modèle reproduit ainsi la loi de Fitts. Il faut noter qu'il existe dans la littérature de nombreux modèles de la loi de Fitts (Meyer et al., 1988; Harris et Wolpert, 1998; Beamish et al., 2006; Tanaka et al., 2006; Guigon et al., 2008b). Il ne s'agit pas ici de proposer un modèle de plus mais de vérifier que la loi de Fitts émerge correctement de notre formalisme dont l'agenda est beaucoup plus large que les questions de variabilité.

La figure 4.6.B met en évidence le fait que tous les paramètres permettent de réguler la précision de l'action et que le modèle ne peut dissocier leurs rôles respectifs dans le contrôle de la variabilité du comportement. Nous ne pouvons donc à ce niveau proposer une interprétation définitive de la façon dont notre formalisme peut aborder le problème de la précision du mouvement, et nous ne pouvons expliquer plus particulièrement comment la description d'une tâche motrice peut être formelle-

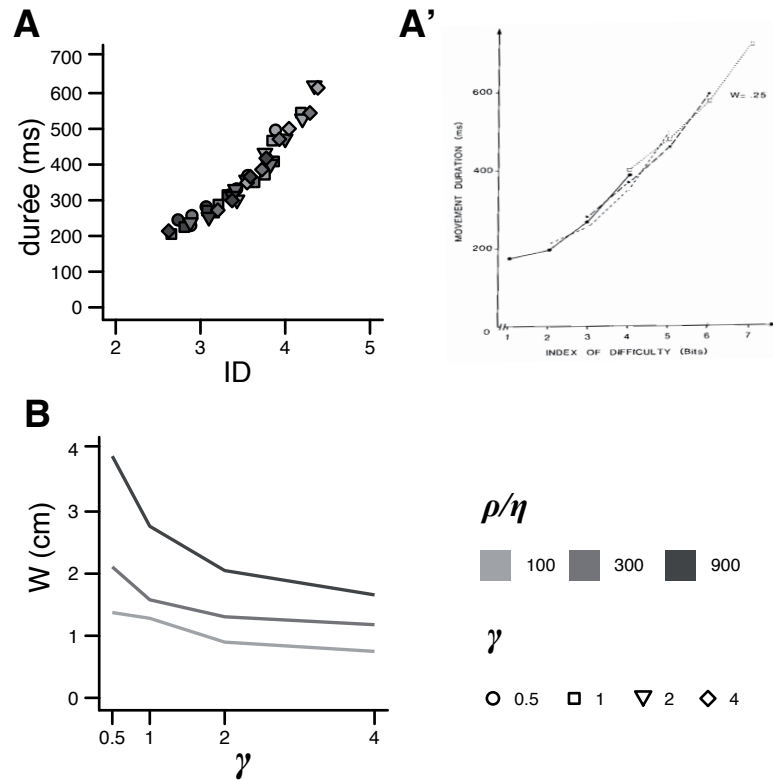


FIGURE 4.6 – *Modulation de la précision*. **A** Durée des mouvements en fonction de leur difficulté **A'** Exemple de la variation de la durée de mouvement pour des difficultés croissantes (amplitude constante), (Jeannerod (1988) à partir des données de Fitts (1954)) **B** Si la durée et la variabilité dépendent de l'amplitude du mouvement pour un ensemble de paramètres internes donnés (voir chapitre 4), la vigueur et le décompte affectent également la durée de mouvement et par conséquent la variabilité terminale.

ment traduite en une valeur précise de paramètres internes. Autrement dit, il reste à expliquer comment, face à une tâche précise (spécifiée par exemple par une taille de cible et des instructions données par l'expérimentateur), le modèle doit déterminer un certain compromis entre vitesse et précision par le biais des paramètres internes ρ , η et γ .

Une autre remarque importante est qu'il existe des cas où les mouvements peuvent être à la fois plus rapides *et* plus précis, en contradiction avec la loi de Fitts. Ce phénomène a par exemple été enregistré chez le singe qui produit des saccades oculaires plus rapides et précises si elles sont associées à une récompense (Takikawa et al., 2002).

Il est toutefois possible d'expliquer ces données en envisageant qu'une augmentation de la récompense permette le recrutement de ressources supplémentaires capables de diminuer le bruit qui corrompt le mouvement. La co-contraction musculaire pourrait tout particulièrement participer à cet effet, comme le suggèrent différentes expériences démontrant un lien fort entre précision du mouvement et régulation fine de la co-contraction (Gribble et al., 2003; Wong et al., 2009; Osu et al., 2009; Selen et al., 2009). Selon notre hypothèse, le mouvement émerge du meilleur compromis entre effort et récompense; la co-contraction étant un effort comme un autre, nous pouvons envisager une extension directe de notre modèle où la co-contraction serait optimisée pour l'avantage qu'elle procure pour atteindre efficacement l'objectif. Cette idée permettrait de donner une explication normative de la co-contraction en tant que planification optimale de l'effort en fonction du but visé. Ce principe est toutefois difficile à tester car il n'existe pas encore de consensus clair sur la causalité physiologique entre co-contraction et variabilité motrice, même si des pistes ont été proposées en ce sens (Selen et al., 2005).

4.4 Flexibilité du mouvement face aux perturbations

Alors que les modèles classiques du contrôle moteur ne produisent que des actions d'une durée définie à l'avance, le formalisme proposé ajuste dynamiquement la durée de l'action jusqu'à ce que l'objectif soit atteint. En effet, le contrôleur re-planifie à tout instant une nouvelle politique d'action qui repose uniquement sur le compromis entre l'effort restant à fournir et le bénéfice escompté au terme du mouvement. Plus précisément, le processus de rétro-contrôle ne tient pas compte du temps qui passe et produit un contrôle qui n'est défini que par ce qu'il reste à faire pour obtenir la récompense visée. La durée d'une action émerge ainsi de l'interaction continue entre ce que le contrôleur anticipe et l'apport réel du mouvement à l'accomplissement du

comportement. Cette flexibilité permet au modèle d'exécuter avec succès des mouvements en présence de perturbations importantes qui ne peuvent être compensées en respectant le temps initialement prévu pour l'action.

4.4.1 Déplacements de la cible

Aussi précise que soit notre capacité à localiser les récompenses dans l'environnement, nos objectifs peuvent évoluer à tout instant sans qu'il soit possible de l'anticiper. La cible d'une action peut par exemple se déplacer de façon imprévue alors même que le mouvement pour l'atteindre a déjà été initié (dans le cas d'une proie pourchassée par exemple). Le comportement généralement observé dans ce cas est une persévération de l'action, c'est-à-dire une poursuite de la cible qui ne s'arrête qu'une fois l'objectif atteint. L'action est alors un long mouvement qui se déroule de façon fluide et ce malgré tous les détours que peut entreprendre une cible en déplacement.

Ce phénomène peut être testé expérimentalement, à moindre échelle, en déplaçant de façon imprévisible une cible visuelle qu'un sujet a pour tâche de toucher avec sa main. Liu et Todorov (2007) par exemple ont utilisé ce type de protocole pour montrer que les sujets sont capables de corriger leurs mouvements sans difficulté pour atteindre une cible changeant brutalement de position au cours du mouvement de pointage. Les auteurs ont également observé que la durée des mouvements dépendait fortement de l'instant de la perturbation : plus le saut de la cible était tardif, c'est-à-dire plus l'exécution du mouvement était avancée, plus les sujets prenaient de temps pour corriger leurs mouvements. Nous avons simulé cette expérience en changeant simplement la position de la cible (\mathbf{x}^* dans l'équation 2.5) à différents instants du mouvement à l'identique de l'expérience de Liu et Todorov (2007). La figure 4.7 montre que le contrôleur ajuste alors sa politique d'action en ligne pour produire des mouvements très similaires à ceux trouvés expérimentalement. La différence de durée entre les différents mouvements s'explique dans nos simulations par le fait que les

corrections tardives sont plus coûteuses énergétiquement et donc sont planifiées avec une durée plus longue par le processus de contrôle en ligne optimal.

Il est intéressant de noter que Liu et Todorov (2007) proposent également un modèle pour expliquer leurs observations. Leur modèle, comme le nôtre, repose sur un principe de contrôle en ligne optimal. Dans leur approche, toutefois, le temps de mouvement n'est pas une propriété émergente du modèle. Ils utilisent en effet les temps de mouvement empiriques pour effectuer leurs simulations, et n'offrent donc pas une vision normative du processus de correction en ligne.

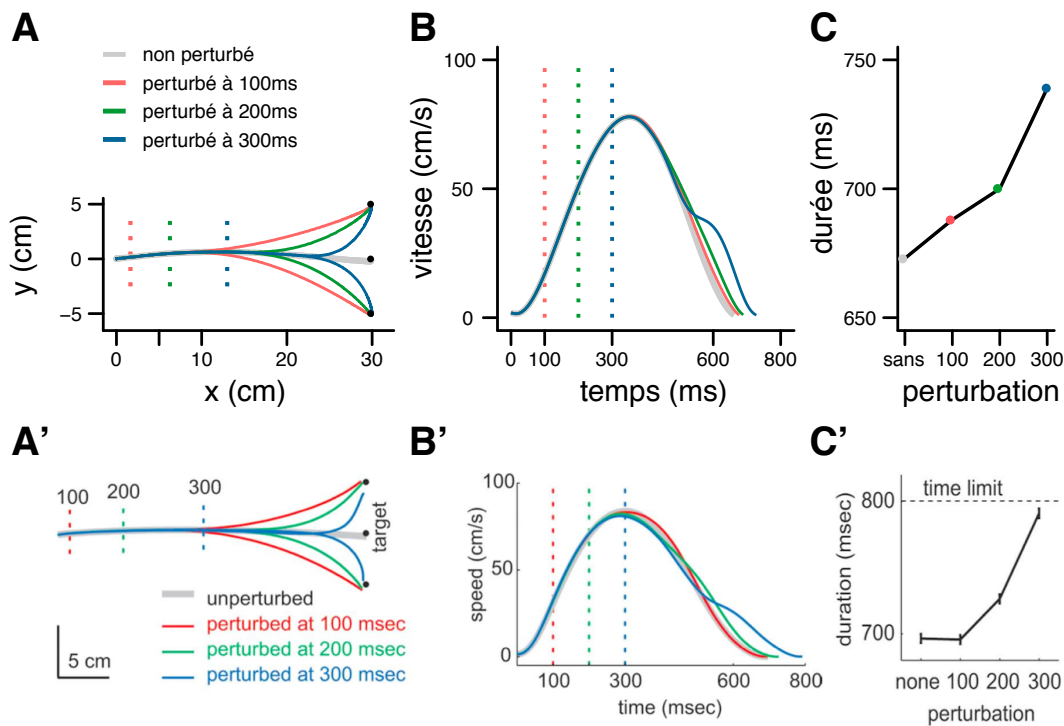


FIGURE 4.7 – *Simulation de l'expérience de saut de cible de (Liu et Todorov, 2007)*
A Trajectoires de pointage vers une cible située à 30 cm qui saute de façon inattendue 5 cm perpendiculairement au mouvement 100 ms (rouge), 200 ms (bleu) ou 300 ms (vert) après le début du mouvement. **B** Profils de vitesse correspondants. **C** Temps d'arrivée à la cible en fonction de la perturbation. Les figures **A'**, **B'** et **C'** sont les données expérimentales correspondantes (Liu et Todorov, 2007). Résultats obtenus avec le bras articulé (f_{III}). Position initiale : (15° , 120°). Même paramètres que sur la figure 4.2.

Les valeurs des paramètres (ρ/η , γ) ont été estimées pour reproduire la durée

de mouvement en l'absence de perturbations. Les trajectoires perturbées sont simulées avec les mêmes paramètres et la flexibilité du mouvement émerge purement du processus de contrôle en ligne. Plusieurs ensembles de paramètres permettent de produire des mouvements de durées identiques, toutefois ces différentes variantes n'occasionnent pas de différences notables sur les résultats présentés.

4.4.2 Altération de la dynamique du corps

Afin de planifier une politique de contrôle, le processus d'optimisation doit connaître la dynamique du système contrôlé (*i.e.* du corps). Cette connaissance peut toutefois être imparfaite ou erronée, lorsque nous utilisons un nouvel outil par exemple ou lorsque des changements physiologiques affectent le système moteur de façon périphérique sans que le cerveau ne puisse les détecter directement. Dans ce cas, les commandes envoyées aux muscles ne guideront généralement pas les membres vers la récompense comme prévu.

Diverses expériences ont modifié artificiellement la dynamique du bras de sujets humains pour observer leur comportement moteur. Un protocole usuel est d'appliquer un champ de force sur la main à l'aide d'un bras robotisé. Dans l'expérience de Shadmehr et Mussa-Ivaldi (1994) par exemple, une force dépendante de la vitesse du mouvement est appliquée sur la main des sujets dans une tâche de pointage dans le plan horizontal. Si, par apprentissage, le contrôleur semble capable de s'adapter à la nouvelle dynamique et de ré-optimiser le contrôle du mouvement (Izawa et al., 2008), les sujets produisent lors de leurs premiers essais des mouvements assez chaotiques : les trajectoires ont une forme de crochet et les profils de vitesse sont plus longs et irréguliers que les profils des mouvements non perturbés. Ce phénomène peut être attribué à une mauvaise configuration du contrôleur qui n'a pas encore adapté sa politique de contrôle pour compenser efficacement (au regard du critère d'optimalité) les perturbations mécaniques. La théorie du contrôle optimal classique, à cause de

son approche rigide de la durée de mouvement, ne possède pas la flexibilité nécessaire pour rendre compte de ces données : à la fin du temps initialement prévu pour l'exécution d'un mouvement non perturbé, l'action s'interrompt brutalement avant d'avoir pu compenser la déviation due au champ de force.

Dans le modèle proposé, le processus de contrôle en ligne permet de corriger le mouvement tant que la cible n'est pas atteinte grâce à une re-planification dynamique du temps de mouvement. Nous avons testé la capacité de notre modèle à expliquer le comportement moteur soumis à un champ de force de la façon suivante. Tout d'abord le contrôleur est optimisé en utilisant la dynamique du bras (f^c) non altérée. Le contrôle fourni sert ensuite à simuler l'évolution du membre dont la dynamique réelle (f^r) est modifiée par l'application du champ de force \mathcal{D} (cf. chapitre 3.3). L'état du bras est alors estimé à l'aide d'une représentation (f^e) de la dynamique du bras, et le contrôleur reprend la main pour diriger le mouvement.

Les résultats de simulation pour l'expérience de Shadmehr et Mussa-Ivaldi (1994) sont montrés sur la figure 4.8. En l'absence de champ de force, les trajectoires de la main sont rectilignes avec un profil de vitesse en cloche, comme vu précédemment (figure 4.8.A et 4.2.B). Si maintenant le même contrôleur est utilisé pour diriger le bras perturbé par le champ de force, les trajectoires produites sont très courbées (figure 4.8.B) et ont des profils de vitesses très irréguliers (figure 4.8.C) quantitativement similaires à ceux mesurés par Shadmehr et Mussa-Ivaldi (1994).

Là encore, un unique ensemble de paramètres estimé pour reproduire les trajectoires non perturbées explique également les trajectoires en présence du champ de force. Le même processus de contrôle optimal en ligne produit indifféremment les mouvements uniformes de la figure 4.8.A et les actions plus alambiquées des figures 4.8.B et C.

Les résultats présentés ont été obtenus avec un estimateur adapté à la perturbation ($f^e = f^r$). Si le processus d'estimation semble apprendre très rapidement (en

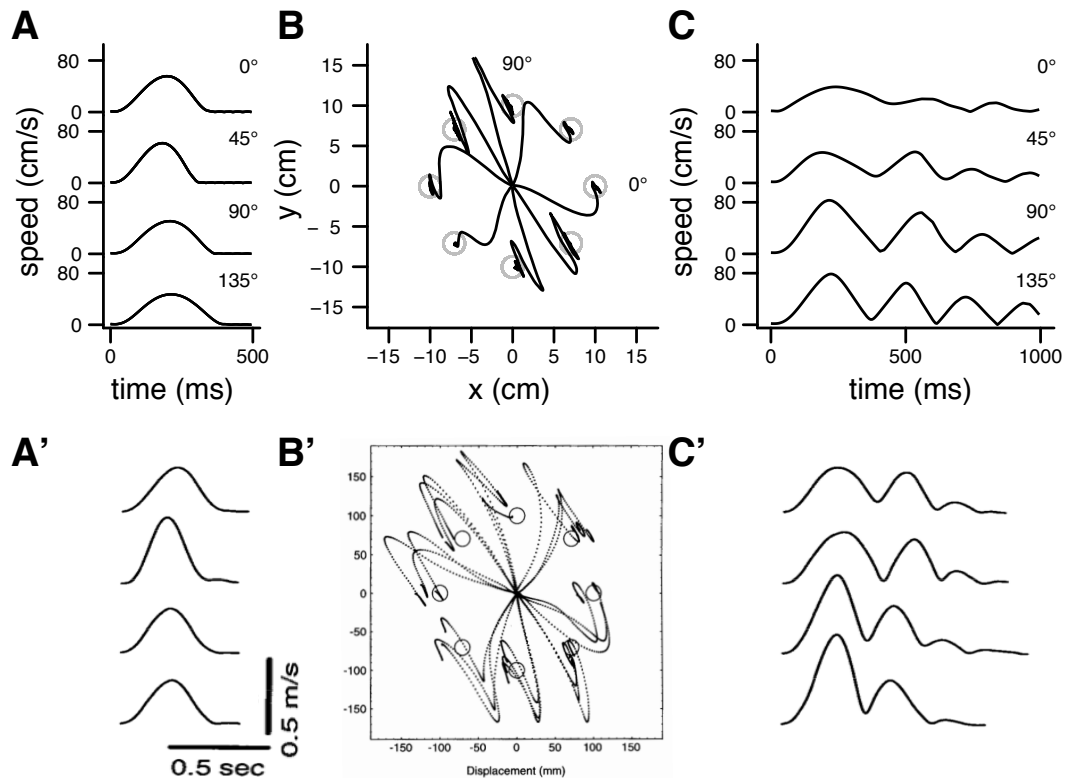


FIGURE 4.8 – *Simulation de l'expérience de champ de force de Shadmehr et Mussa-Ivaldi (1994)* **A** Profils de vitesse des mouvements dans quatre directions en l'absence de champ de force ($f^c = f^r$). Les trajectoires (non montrées) sont qualitativement équivalentes à celles de la figure 4.2.B. **B** Trajectoires de la main lors de la première exposition au champ de force ($f^c \neq f^r$). **C** Profils de vitesse des mouvements perturbés montrés en B pour les mêmes directions que A. Résultats obtenus avec le bras articulé (f_{III}). Position initiale : $(15^\circ, 100^\circ)$. Champ de force : $\mathcal{D} = [-10.1 \ -11.2; -11.2 \ 11.1]$. Même paramètres que sur la figure 4.7. **A'**, **B'** et **C'** Données expérimentales correspondantes (Shadmehr et Mussa-Ivaldi, 1994).

quelques essais) à anticiper les changements de dynamique (Flanagan et al., 2003), bien avant que le contrôleur s'adapte lui aussi aux perturbations, les tous premiers mouvements perturbés devraient être simulés avec un estimateur naïf ($f^e = f^c \neq f^r$). Dans ce cas, les simulations que nous avons réalisées présentent des déviations plus importantes que ce qui est observé expérimentalement. Si les mouvements arrivaient tous en fin de compte à toucher la cible, le comportement moteur modélisé était quantitativement moins bon qu'espéré. Cette différence peut s'expliquer par l'absence dans notre modèle des réflexes et de la co-contraction musculaires, deux mécanismes qui,

à l'instar des propriétés musculaires décrites chapitre 3.3, favorisent la stabilité des mouvements en augmentant la résistance du corps aux perturbations. Les réflexes pourraient être facilement intégrés au formalisme proposé (sous la forme d'un rétro-contrôle négatif proportionnel à l'écart entre l'état réel des muscles et celui prévu par le cerveau, par exemple), cependant cette démarche complexifie notablement la dynamique du système avec des paramètres dont il est délicat d'estimer la valeur à partir des données expérimentales. La co-contraction est plus difficile à incorporer dans notre modèle. Si la causalité physiologique entre la co-contraction musculaire et l'augmentation de la résistivité mécanique d'une articulation peut tout à fait être formalisée, il est par contre difficile d'expliquer *pourquoi* le cerveau opte pour une stratégie de co-contraction particulière. La co-contraction nécessite en effet plus de contrôle, un effort supplémentaire que le processus d'optimisation tend à réduire par définition. Une solution possible est d'explicitier en quoi l'effort investi dans la co-contraction peut augmenter en contrepartie les bénéfices de l'action (en diminuant le temps de mouvement anticipé par exemple). Nous pourrions alors imaginer que l'effort d'une action intentionnelle se décompose optimalement en un effort de mouvement et un effort de co-contraction (Burdet et al., 2001). Ces remarques sur la participation des réflexes et de la co-contraction à la formation des comportements moteurs ne remettent pas en cause les résultats de notre modèle : ces mécanismes ne se substituent pas au processus de contrôle en ligne mais le complètent en stabilisant (activement ou passivement) la dynamique du système face aux perturbations mécaniques.

Le contrôle optimal en ligne proposé permet d'expliquer avec succès le comportement moteur lors de perturbations logistiques ou mécaniques. Quels que soient les événements qui menacent l'accomplissement d'une action (bruit, erreur dans la localisation de la cible, mauvaise représentation de la dynamique), notre modèle est

capable de produire une réponse optimale pour regagner la récompense visée grâce à un processus uniforme et autonome de contrôle prédictif. Là encore, il existe de nombreux modèles qui tentent d'expliquer les caractéristiques des mouvements dans les champs de force (Shadmehr et Mussa-Ivaldi, 1994; Bhushan et Shadmehr, 1999; Sanner et Kosha, 1999; Burdet et al., 2006). Cependant, aucun de ces modèles ne peut rendre compte de façon normative de la durée des mouvements perturbés. L'allongement des temps de mouvement dans ces modèles s'explique par la mise en jeu d'une rétro-action liée aux propriétés viscoélastiques du bras. Même si cette explication peut sembler raisonnable dans ce cas, elle ne permet pas de rendre compte des observations liées aux sauts de cibles (voir plus haut ; Liu et Todorov, 2007).

L'organisation du comportement

Le chapitre précédent s'est focalisé sur les aptitudes de notre modèle à expliquer la coordination motrice de façon réaliste. D'un autre côté, le formalisme proposé permet d'aborder le comportement dans ses aspects décisionnels, c'est-à-dire d'étudier le comportement d'un animal face à un problème de choix entre plusieurs options possibles, soit plusieurs actions menant à des récompenses différentes. En effet, le modèle permet de décrire une action non seulement au travers de son contenu (le contrôle que demande son exécution) mais également d'une manière plus globale en ne considérant que sa valeur comportementale subjective (son utilité). Grâce à cette dualité, les actions peuvent alors être manipulées comme des briques élémentaires de comportement qui peuvent être comparées et combinées pour répondre à des tâches complexes, une description de l'action très puissante pour expliquer le comportement à un niveau macroscopique (Sutton et Barto, 1998; Daw et Doya, 2006; Rangel et al., 2008). En faisant tomber la barrière entre optimisation des récompenses et gestion des efforts, le formalisme proposé suggère que derrière cette description à un niveau macroscopique du comportement se cache des interactions riches entre sélection et exécution d'une action. L'intégration précise du coût de l'action lors de son évaluation permet d'attaquer le problème posé par les tâches de décision en situation d'effort. Inversement, notre modèle prédit des effets du contexte comportemental sur le contenu des actions et donne sens à des interactions entre la valeur des objectifs visés et la régulation de la réalisation motrice. Les résultats ci-dessous soulignent ainsi la faculté du modèle à reproduire la perméabilité réciproque des processus de décision et de

coordination motrice à la motivation et à l'effort.

5.1 Effort et décision

Le modèle que nous avons présenté attribue une utilité à une action potentielle non seulement en se basant sur la récompense qu'elle permet d'obtenir mais également en intégrant de façon précise le coût de sa réalisation. Contrairement aux modèles classiques de sélection de l'action, nous pouvons ainsi faire des prédictions quantitatives sur le choix d'un animal dans une tâche où gains et efforts contribuent concurremment à la décision.

5.1.1 Procédure de décision

Si plusieurs récompenses sont accessibles simultanément, l'utilité optimale J^* d'une action (solution de l'équation 2.9) permet de classer les options disponibles par ordre de leur valeur comportementale nette (récompense moins effort). Ce classement permet de définir une préférence parmi les alternatives possibles et ainsi de guider le choix d'action sur un critère écologique.

Intuitivement, le choix le plus intéressant est celui de l'action de plus grande utilité, celle qui offre le meilleur rapport gains sur investissement énergétique. L'étude du comportement animal montre pourtant que la stratégie la plus immédiate qui consiste à toujours choisir l'action d'utilité maximale n'est pas celle suivie de façon générale. De nombreuses expériences indiquent que la stratégie la plus répandue est plutôt de choisir les actions à une fréquence relative à leur utilité respective. Ainsi, l'action de plus grande utilité sera sélectionnée d'autant moins souvent que les autres possibilités auront une utilité importante. Si deux actions ont la même utilité, elles auront la même probabilité d'être choisies. Ce phénomène, qui contrevient au principe du maximum d'utilité, est connu sous le nom de *loi d'assortiment* (Sugrue et al., 2005). De

nombreuses études on tenté d'expliquer cette anomalie (comme le résultat d'un bruit dans le processus de décision par exemple) ou de la rationaliser (mécanisme d'exploration, optimisation des accès aux ressources en cas de compétition avec d'autres animaux) sans toutefois convenir d'une interprétation unique. Nous ne chercherons pas à expliciter plus en avant ce phénomène et pour les besoins de la démonstration nous nous contenterons de le capturer de manière purement descriptive.

Une méthode répandue pour modéliser la loi d'assortiment est d'utiliser une fonction soft-max (McClure et al., 2003; Daw et Doya, 2006). Prenons par exemple une situation où deux récompenses (r_1 et r_2) sont présentes dans l'environnement. Pour chacune d'elles, il est possible de planifier un mouvement optimal et par conséquent de définir une unique utilité d'action (respectivement J_1 et J_2). La probabilité P_i de choisir l'action i dirigée vers la récompense r_i est alors donnée par :

$$P_i = \frac{e^{\beta J_i}}{e^{\beta J_1} + e^{\beta J_2}} \quad (5.1)$$

D'après cette équation, plus une action aura une utilité relative importante, plus elle aura de chances d'être choisie. Le paramètre β permet d'ajuster le niveau de rigueur du processus de décision : un petit β rend la décision aléatoire, un très grand β garantit que l'action de plus grande utilité est toujours choisie.

5.1.2 Situation de conflit entre effort et récompense

Si l'effort associé à l'obtention d'une récompense est négligeable, l'utilité d'action que nous calculons est équivalente à l'utilité anticipée de la récompense et la règle de décision décrite ci-dessus se comporte comme une règle de décision basée uniquement sur la valeur de récompense. Notre formalisme se réduit alors à l'approche classique de la théorie du renforcement et conserve toute sa puissance explicative du comportement. Dans le cas maintenant où les récompenses ne sont pas toutes éga-

lement accessibles mais demandent des efforts différents, à cause de leurs distances respectives ou d'obstacles bloquant spécifiquement une ressource par exemple, notre modèle fait des prédictions différentes. En effet, l'utilité d'action telle que nous la définissons déduit le coût de l'action du bénéfice de la récompense, ce qui implique qu'une récompense semblera d'autant moins intéressante qu'elle est difficile d'accès. Le modèle est donc adapté pour étudier les tâches de décision où les récompenses et les efforts entrent en conflit dans l'évaluation des actions.

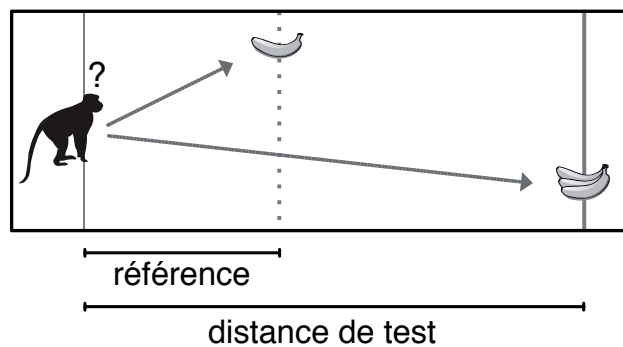


FIGURE 5.1 – *Exemple de situation de conflit entre coût et bénéfices.* Schéma inspiré du protocole de Stevens et al. (2005).

Dans le protocole utilisé par Stevens et al. (2005) par exemple, schématisé sur la figure 5.1, un singe doit choisir entre récupérer une petite récompense localisée à une courte distance (référence) ou bien obtenir une récompense plus conséquente mais située plus loin (test). Dans cette situation, la plus grande récompense n'est pas toujours préférable car il faut fournir un plus grand effort pour l'obtenir.

Nous avons modélisé chaque singe comme une simple masse (objet I). Ensuite, nous avons calculé l'utilité (équation 2.8) pour cet animal virtuel d'aller chercher une récompense simple ($r = 1$) localisée à une faible distance (\mathbf{x}^* à 35 cm) ou une récompense plus grande ($r = 3$) à une distance éventuellement plus grande (\mathbf{x}^* entre 35 et 245 cm). La figure 5.2 montre les résultats de la simulation. Si les deux cibles sont à la plus faible distance, l'option la plus récompensée a toujours la plus grande utilité, et sera donc toujours préférée. L'utilité de l'option la plus récompensée diminue

toutefois avec la distance, à cause de l'effort de plus en plus grand qu'elle demande. Au delà d'une certaine distance (ici quand x^* est à plus de 105 cm), attraper la grande récompense devient trop coûteux et moins rentable que la solution de facilité. Pour une distance de test bien déterminée, le rapport bénéfice/coût est identique pour les deux options qui ont alors la même utilité : c'est le point d'indifférence, c'est-à-dire la situation où l'animal ne préfère aucune option en particulier.

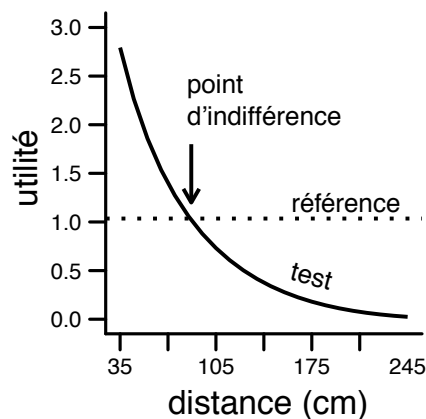


FIGURE 5.2 – *Utilité d'une action en fonction de la distance et de la taille de la récompense.* Utilité de l'option de référence (*distance* = 35 cm, $r = 1$; pointillé) et de l'option de test (*distance* = de 35 à 245 cm, $r = 3$; plein). Le point d'indifférence, où les deux options ont une utilité égale, est indiqué par une flèche. Simulé avec l'objet I pour $\rho/\eta = 1$ et $\gamma = 2$.

Ce résultat est un exemple de la capacité beaucoup plus générale du modèle à expliquer l'effet de la difficulté de l'action sur la décision qui a largement été rapporté dans la littérature (Salamone et al., 2003; Stevens et al., 2005; Walton et al., 2006; Rudebeck et al., 2006; Croxson et al., 2009; Kurniawan et al., 2011). Cette propriété vient de notre définition du comportement comme la maximisation du compromis entre le bénéfice et le coût de l'action (équation 2.8).

La force du modèle est que l'évaluation de l'effort associé à l'action repose sur une planification précise du coût moteur, ce qui garantit une cohérence naturelle entre les coûts anticipés lors de la décision et les coûts réellement investis dans l'exécution motrice. Le modèle permet ainsi d'expliquer les données expérimentales dans des tâches

de décision où l'anticipation des coûts moteurs repose sur une parfaite évaluation de la dynamique corporelle (Cos et al., 2011).

5.1.3 Une représentation commune de l'effort

L'évaluation du compromis entre effort et récompense est modulée, dans notre formalisme, par un ensemble de paramètres internes qui pondèrent la sensibilité à la récompense (ρ), à l'effort (η) et au temps (γ). Comme la résolution de ce compromis est responsable à la fois de l'émergence de la durée des mouvements (équation 2.13) et du choix en situation d'effort (nous venons de le voir), la valeur de ces paramètres doit influencer simultanément le comportement moteur et la décision.

Si cette hypothèse est vraie, nous devrions être capable de prédire le choix en situation d'effort uniquement en analysant la durée des mouvements. Pour le tester, nous avons étudié l'expérience de Stevens et al. (2005) qui rapporte justement la vitesse de mouvement et le choix en situation d'effort pour deux espèces de singes, les tamarins et les marmousets.

Nous avons tout d'abord, pour chaque espèce, établi un « profil » de la gestion de l'effort en estimant les paramètres internes capables d'expliquer la vitesse de mouvement. Pour cela, nous avons modélisé des singes comme de simples masses qui se déplacent. Chacun de ces singes virtuels est caractérisé par ses paramètres internes et donc par un comportement moteur qui lui est propre. Stevens et al. (2005) (figure 5.3.A) rapporte pour chaque espèce de singe la durée de déplacement pour l'option de référence (plus petite distance : \mathbf{x}^* à 35 cm, petite récompense : $r = 1$) et l'option la plus difficile (distance maximale : \mathbf{x}^* à 275 cm, grande récompense : $r = 3$). Nous avons alors sélectionné parmi nos individus virtuels deux sous-populations, une pour chaque espèce de singe, de telle sorte que la durée des déplacements de chaque groupe soit statistiquement identique aux durées rapportées par Stevens et al. (2005). La figure 5.3.B montre que cette procédure permet effectivement d'identifier deux

profils différents, et plus particulièrement que les marmousets sont globalement plus sensibles à l'effort que les tamarins.

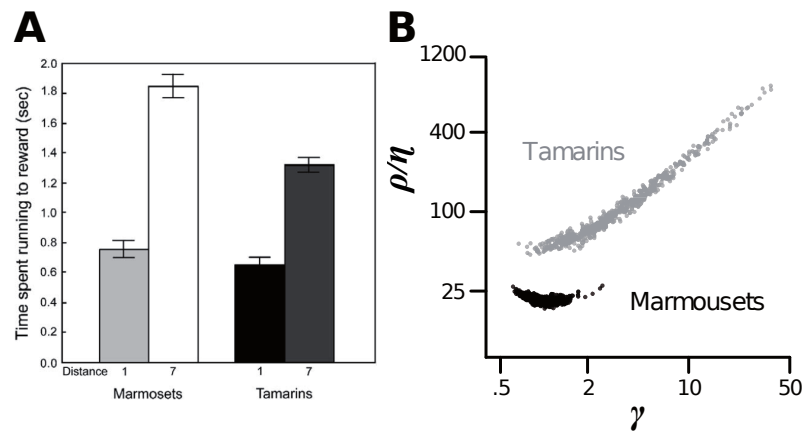


FIGURE 5.3 – *Estimation des paramètres internes à partir de la durée des mouvements.* **A** Vitesse de déplacement en fonction de l'espèce et de la distance de la récompense (figure 3 de Stevens et al. (2005)). **B** Représentation des deux populations de singes virtuels en fonction des paramètres internes qui les caractérisent. Chaque point correspond à un individu.

A partir de là, nous avons simulé le comportement de chaque population de singes dans la tâche de choix utilisée par Stevens et al. (2005) et que nous avons décrite dans la section précédente (voir figure 5.1). Pour y arriver, nous avons calculé comme précédemment l'utilité de chaque option, mais cette fois-ci en pondérant le compromis entre effort et bénéfice selon le profil caractéristique de chaque individu. La figure figure 5.4.A montre les résultats pour un individu typique de chaque population (dont les paramètres internes ont pour valeur la moyenne de ceux de son groupe). Il apparaît que la sensibilité particulière à l'effort des marmousets se traduit par un point d'indifférence à bien plus faible distance (environ $2m$) que pour les tamarins (supérieure à $3m$).

En appliquant maintenant la règle de soft-max pour simuler la décision de chaque population de singes, nous pouvons établir la probabilité qu'un animal choisisse d'aller à la plus grande récompense en fonction de la distance à laquelle elle se trouve. Ces simulations font apparaître que nos marmousets sont très sensibles à la distance et

délaissent volontiers la grande récompense si celle-ci se trouve trop loin (au delà de 175 cm) alors que les tamarins choisissent presque toujours la récompense la plus grande quel que soit l'effort nécessaire pour l'atteindre (figure 5.4), exactement comme mesuré expérimentalement par Stevens et al. (2005).

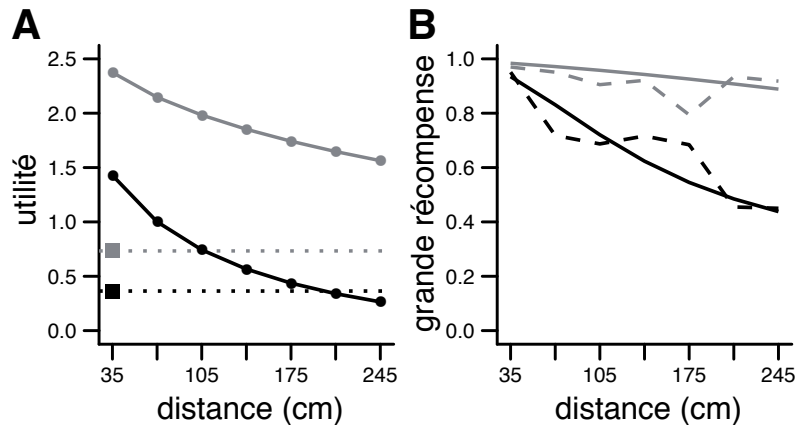


FIGURE 5.4 – Résultats de la simulation de l'expérience de Stevens et al. (2005). **A** Utilité de l'option de référence (carrés, $r = 1$) et de l'option de test (ronds, $r = 3$) calculés pour les marmousets (noir) et les tamarins (gris). Les paramètres sont ceux de l'individu moyen pour chaque espèce, soit $\rho/\eta = 22$ et $\gamma = 1$ pour les marmousets, $\rho/\eta = 90$ et $\gamma = 2.7$ pour les tamarins. Résultats obtenus avec le point masse simple. **B** Probabilité de choisir la plus grosse récompense en fonction de sa distance pour les deux espèces de singe (mêmes couleurs que **A**). Les lignes pleines sont les prédictions du modèle, les lignes pointillées sont les données de Stevens et al. (2005).

La température β est un paramètre libre qui est estimé empiriquement pour correspondre aux données expérimentales (ici, $\beta = 2.5$). Si la valeur de β affecte quantitativement nos résultats, elle ne peut remettre en cause le fait que les marmousets présentent un changement de préférence vers 175 cm alors que les tamarins sont relativement indifférents à la distance de la récompense, une caractéristique due uniquement aux paramètres internes du modèle (figure 5.4.A).

En conclusion, ces résultats sont en faveur de la théorie selon laquelle les comportements moteurs et la décision sont sous-tendus par des facteurs communs. Cette hypothèse appelle à de plus amples investigations, une tâche toutefois difficile à l'heure actuelle à cause du manque de littérature empirique reportant simultanément le com-

portement moteur et les données en décision de façon quantitative. Cependant, nous verrons dans le chapitre suivant que des arguments physiologiques concourent également à l'idée d'un socle commun à la régulation motrice et décisionnelle.

5.2 Sensibilité du mouvement au contexte comportemental

Notre modèle permet de rendre compte de la perméabilité de la décision au coût réel des actions. Réciproquement, il prédit que le contenu de l'action, c'est-à-dire le mouvement, est dépendant de la valeur de l'objectif visé. L'effort investi dans un mouvement peut se mesurer à sa vitesse d'exécution : une action plus rapide demande un recrutement plus important des muscles, occasionne une plus grande dépense énergétique et nécessite des commandes motrices de plus grande amplitude. La vitesse de mouvement est donc une variable primordiale du comportement en tant que mesure indirecte de l'effort investi dans une tâche.

Les modèles existants du contrôle moteur postulent que le processus de planification et d'exécution du mouvement suit un principe d'économie qui lui est propre, indépendant de la valeur de l'objectif visé. Cette herméticité aux facteurs motivationnels se trahit dans ces formalismes par une incapacité à expliquer comment est définie la durée de mouvement et donc comment est modulé l'effort d'une action.

5.2.1 Motivation et vitesse

Dans notre modèle, la durée de mouvement n'est pas un paramètre à déterminer mais une propriété émergente du processus d'optimisation. Plus exactement, le temps de mouvement est sélectionné pour maximiser le compromis entre la récompense décomptée et l'effort. Par définition, il est donc sensible à la motivation. Plus précisément, plus la récompense sera grande, plus grand sera l'effort optimal et plus

rapide sera le mouvement (cf. chapitre 2). Ainsi, notre modèle donne une interprétation normative et intuitive aux nombreuses données empiriques qui rapportent un effet de la valeur de l'objectif (la taille de la récompense par exemple) sur la vitesse du mouvement : les rats courent d'autant plus vite (Crespi, 1942; Peciña et al., 2003), les singes (Takikawa et al., 2002; Watanabe et al., 2003; Kobayashi et al., 2006) et les humains (Xu-Wilson et al., 2009) font des saccades d'autant plus rapides que l'objectif visé est associé à une récompense de grande valeur. Plus généralement, notre formalisme explique pourquoi la motivation influence directement l'effort investi dans nos actions (Berridge et Robinson, 1998; McClure et al., 2003; Pessiglione et al., 2007; Schmidt et al., 2008; Berridge et Aldridge, 2009). Il est important de noter que cette modulation de la vitesse que le modèle reproduit n'est pas liée à une causalité instrumentale entre la quantité d'effort fournie et la récompense effectivement obtenue : aller plus vite ne fait pas gagner plus. L'augmentation de l'effort avec la motivation émerge dans notre modèle de l'optimisation de la valeur subjective des actions (équation 2.13), un critère certes normatif mais dont l'avantage écologique reste spéculatif à ce niveau de description.

5.2.2 Le cas des séquences motrices

Si la motivation influence directement la durée des mouvements les plus simples, notre modèle prédit un effet plus indirect sur la vitesse des actions réalisées en séquence. Pour comprendre cette propriété, il nous faut d'abord expliciter comment rendre compte de la formation de séquences comportementales dans notre formalisme. Prenons l'exemple classique d'un labyrinthe en T dans lequel un rat doit se déplacer pour récupérer de la nourriture, comme illustré sur la figure 5.5.

Si le rat se trouve au point \mathbf{x}^\times , il est face à un problème de décision comme décrit dans le chapitre précédent. Dans notre situation, l'animal va évidemment choisir d'aller dans la branche de droite au bout de laquelle (au point \mathbf{x}^*) se trouve une

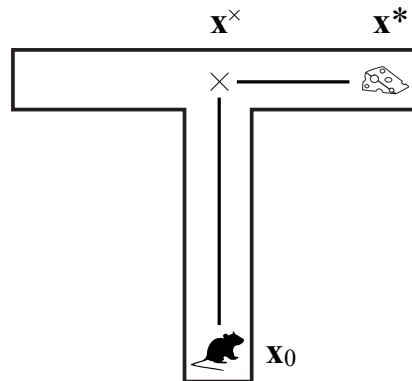


FIGURE 5.5 – *Exemple de la formation d'une séquence comportementale.* Un rat se trouve dans un labyrinthe à la position x_0 . Une récompense r est localisée en x^* . Pour atteindre cette nourriture, l'animal doit d'abord passer par le point x^x où il ne touche aucune récompense.

récompense r . Si maintenant le rat se trouve au point x_0 , il doit d'abord se déplacer en x^x avant de pouvoir s'engager dans la branche récompensée. Qu'est-ce qui motive notre animal à faire ce mouvement puisqu'il n'obtient aucune récompense en x^x ? Dans le formalisme du RL, l'action d'aller de x_0 à x^x est « récompensée » par l'opportunité de pouvoir ensuite effectuer l'action d'aller de x^x à x^* , la valeur subjective de cette opportunité étant mesurée par son utilité. Autrement dit, la valeur de la récompense r se rétro-propage le long de la chaîne comportementale pour motiver les actions successives. Contrairement à la théorie du renforcement qui ne considère les actions que comme des *choix* transparents, nous considérons dans notre modèle qu'une action a un coût (énergétique et temporel) qui influence fortement son utilité. Ainsi, la motivation à effectuer un mouvement décroît avec chaque action qui le sépare encore de l'objectif. En effet, le premier mouvement d'une séquence doit décompter de la récompense qu'il contribue à atteindre le coût de tous les mouvements qu'il reste à accomplir. A cause de la perméabilité inhérente à notre modèle du mouvement à la motivation, ceci implique qu'une même action ne sera pas réalisée à la même vitesse en fonction de si elle est produite isolément ou au sein d'une séquence.

Ce principe pourrait expliquer le phénomène rapporté par la littérature sous le nom d'*avantage de la cible unique*. Ce phénomène se traduit chez les sujets humains par une exécution plus rapide des mouvements atomiques par rapport aux mouvements initiant des séquences. Afin de tester si notre modèle permet effectivement d'expliquer cet effet de séquence sur la production motrice, nous avons simulé l'expérience de Adam et al. (2000) schématisée sur la figure 5.6.A. Dans cette tâche, les sujets doivent réaliser un mouvement suivi éventuellement d'un deuxième d'une amplitude variable. Les mouvements simples ont été simulés en utilisant une seule cible de valeur $r = 1$. Le premier mouvement de chaque séquence a quant à lui été simulé en utilisant une cible intermédiaire, non récompensée, dont la valeur est égale à l'utilité du mouvement vers la cible finale où se trouve effectivement la récompense, *i.e.* $r = J$. Les résultats, montrés sur la figure 5.6.B, reproduisent bien les données empiriques.

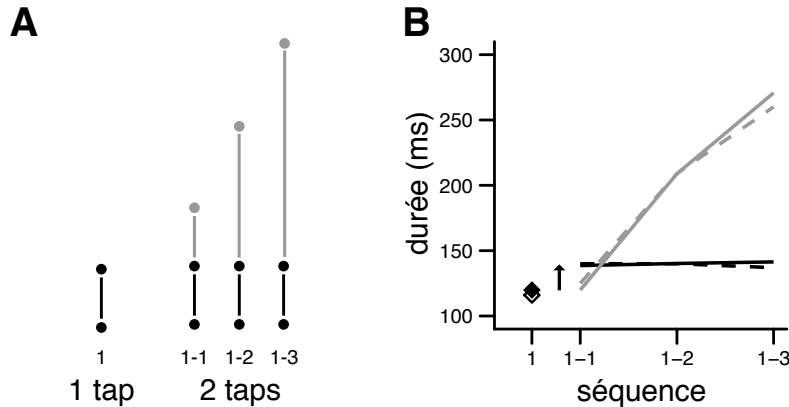


FIGURE 5.6 – *Tâche de production de séquences motrices*. **A** Schéma de l'expérience de Adam et al. (2000). Selon les conditions, le sujet doit effectuer un mouvement simple (1 tap) de 8 cm ou deux mouvements successifs (2 taps) de 8 cm et 8, 16 ou 40 cm respectivement. **B** Durée des différents mouvements. La flèche montre que le premier mouvement (en noir) est plus rapide s'il est réalisé seul (losange) qu'à l'intérieur d'une séquence (ligne). La durée du deuxième mouvement (ligne grise) dépend uniquement de la distance à la cible finale. Résultats de simulations obtenus avec l'objet I avec les paramètres $\rho/\eta = 2700$, $\gamma = 2.5$. Les lignes pointillées et le losange blanc sont les données empiriques de Adam et al. (2000).

Ces résultats montrent dans l'ensemble que l'aspect intentionnel des mouvements

se trahit non seulement dans leur direction (le choix qu'ils réalisent *de facto*) mais également dans leur vitesse d'exécution qui reflète finement la motivation qui a poussé leur préparation. L'effet de séquence suggère que non seulement le contexte environnemental influence l'action mais que toute la structure du comportement se répercute dans la planification de l'action dans ses moindres détails moteurs. Le processus de chaînage des actions à la base des comportements complexes se révèle finalement autant dans le contenu moteur des actions que dans leur description à un niveau macroscopique. Cette hypothèse sous-entend que le degré d'anticipation de séquences d'actions, *i.e.* le nombre de coups prévus à l'avance, devrait pouvoir se mesurer aux fluctuations de la durée des tous premiers mouvements, un effet qu'il serait facile de tester expérimentalement.

Modulation de l'action et pathologies

Nous avons exposé, dans les chapitres précédents, comment les différents principes de notre formalisme font émerger les propriétés fondamentales du comportement dans ses aspects macroscopiques (chapitre 5) comme dans le détail de leur exécution motrice (chapitre 4). Nous avons toutefois négligé de décrire comment le comportement était modulé en dehors des modalités liées à la résolution d'une tâche particulière. Le comportement est en effet soumis à une régulation importante par des facteurs endogènes tels que la motivation ou la fatigue par exemple. Dans notre modèle, plusieurs paramètres (ρ , η et γ) influencent le comportement sans affecter ses propriétés cardinales. Ce chapitre tente de dresser à l'égard de ces paramètres formels une taxonomie des régulations comportementales. Nous ébaucherons également à partir de cette catégorisation un parallèle entre les paramètres du modèle et la neurophysiologie de l'action.

Revenons tout d'abord sur les paramètres du modèle. D'un côté, ρ et η pondèrent respectivement l'influence de la récompense et de l'effort dans l'évaluation subjective des opportunités d'action associées à une tâche donnée. Ces deux facteurs sont en réalité redondants et seul leur rapport compte dans la détermination de l'effort à investir dans un mouvement, et nous avons appelé *vigueur* le ratio ρ/η pour souligner la dimension motivationnelle de l'interaction de ces paramètres. Le facteur de décompte (γ), lui, n'intervient pas directement dans le compromis entre effort et récompense. Son rôle est plutôt de nuancer la planification en dépréciant les prédictions trop éloignées dans le temps et donc plus incertaines. Le paramètre γ est clairement

d'une nature différente de la vigueur. Il possède une dimension temporelle plutôt que motivationnelle. La vigueur et le décompte n'affectent pas les caractéristiques fondamentales de la coordination et du comportement. Les effets de ces paramètres étant de plus qualitativement similaires, il est délicat de démontrer une dichotomie claire entre ces facteurs. Nous allons voir dans ce chapitre que si les effets de ces paramètres sont qualitativement similaires, la vigueur et le décompte semblent toutefois instancier deux modes distincts de régulation du comportement.

6.1 Effet de la vigueur

Nous avons montré dans le chapitre 5 que l'effort requis pour obtenir une récompense influençait fortement le comportement. Plus précisément, au delà d'un seuil critique d'effort à investir, un animal délaisse une grosse récompense pour lui préférer une récompense moins intéressante mais plus facile d'accès. La modulation de la vigueur produit dans notre modèle un déplacement de l'effort critique dans ces tâches de décision. En effet, une modification du ratio ρ/η change le bénéfice relatif d'une récompense par rapport à l'effort qu'elle demande. Lorsque la vigueur augmente, les ressources difficiles à obtenir deviennent plus intéressantes et la décision penche plus facilement vers les options de type gros effort / grande récompense. Inversement, une diminution de la vigueur entraîne une certaine paresse et un biais de la décision vers les solutions moins coûteuses (figure 6.1.A).

La propension à investir plus d'effort dans l'action avec l'augmentation de la vigueur se répercute également dans le contrôle du mouvement. En effet, conférer plus d'énergie au mouvement permet de recruter plus vivement les muscles et ainsi d'aller plus vite vers une cible donnée. Cette accélération générale des mouvements n'affecte pas les propriétés de la coordination, et une augmentation de la vigueur engendre ainsi un déplacement de la loi de mise à l'échelle vers les durées de mouvements plus

courtes (6.1.B).

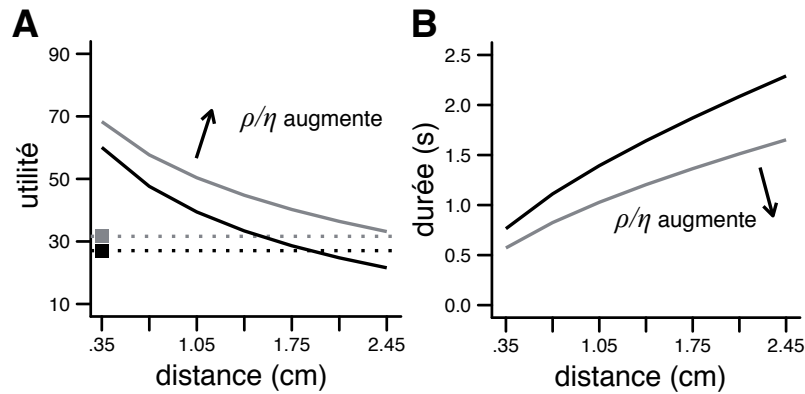


FIGURE 6.1 – *Effet de la vigueur* **A** Tâche de décision en situation d’effort identique à celle décrite figures 5.1 et 5.2. Quand le ratio ρ/η augmente, l’utilité des actions augmentent en faveur des mouvements les mieux récompensés et entraîne un changement du point d’indifférence entre l’option de référence (carré) et l’option de test (ligne). **B** Durée des mouvements pour les différentes distances de test. Paramètres pour les deux figures : $\rho/\eta = 50$ (noir) et $\rho/\eta = 16$ (gris), $r = 1$ et $\gamma = 2$. Résultats obtenus avec le point masse simple.

6.2 Effet du décompte

Le facteur de décompte peut être vu comme une forme de pression temporelle qui pénalise les actions trop longues ou trop lentes. Le relâchement de cette contrainte (une augmentation de γ) entraîne une revalorisation des actions les plus longues (et donc les plus difficiles). Cette réévaluation produit dans les situations d’effort un changement du critère de décision identique à une augmentation de la vigueur (figure 6.2.A). En revanche, un mouvement vers une cible donnée sera plus lent lorsque la récompense est dévaluée moins vite. Autrement dit, l’affaiblissement de la contrainte du temps induit une diminution globale de la vitesse des mouvements (figure 6.2.B).

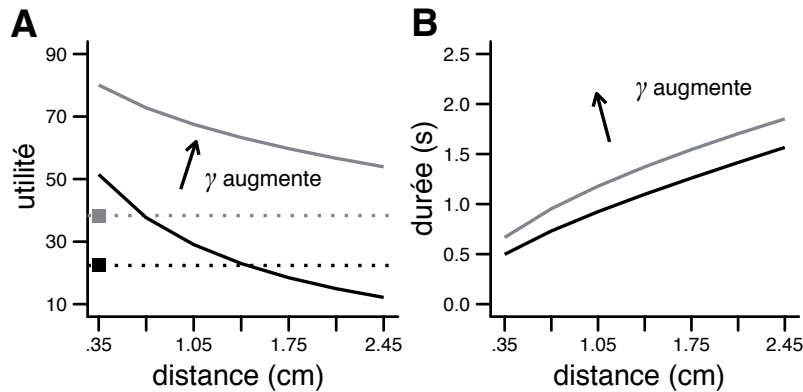


FIGURE 6.2 – *Effet du décompte temporel* Utilité d'action (A) et durée de mouvement (B) en fonction de la distance, à l'identique de la figure 6.1 mais en jouant sur la valeur du décompte. Paramètres : $\rho/\eta = 50$, $r = 1$ et $\gamma = 1$ (noir) ou $\gamma = 4$ (gris).

6.3 Deux modes de régulation

Dans l'ensemble, ces résultats démontrent que la vigueur et le décompte ne peuvent être distingués uniquement sur la base de leur impact sur la décision ou le mouvement individuellement. Toutefois, ces deux paramètres ont des effets différentiels sur le sens dans lequel ils modulent conjointement le critère de décision et la durée des actions motrices. Ainsi, le modèle postule deux modes différents de régulation de l'action. Le premier est d'ordre motivationnel et permet de moduler l'effort à engager dans le comportement (distance à parcourir, vitesse). Le second est quant à lui lié à la dimension temporelle de l'action et met en perspective le risque inhérent au délai qui sépare l'instant de la planification du moment où la récompense doit être obtenue.

6.3.1 L'hypothèse dopaminergique

La régulation motivationnelle par la vigueur ressemble très fortement à l'influence qu'a la dopamine (DA) sur le comportement animal. Chez le rat par exemple, une diminution de la transmission dopaminergique consécutive à l'injection systémique d'antagonistes de la DA occasionne en effet à la fois un abaissement du critère de

décision (Walton et al., 2006; Bardgett et al., 2009) vers les options de moindre effort et en parallèle un ralentissement de la vitesse moyenne de locomotion (Denk et al., 2005; Salamone et al., 2007; Floresco et al., 2008). Inversement, les souris manipulées génétiquement pour produire de la DA en grande quantité ont tendance à exercer plus d'effort et à se déplacer plus vite que les souris normales (Peciña et al., 2003; Cagniard et al., 2006), de même que les souris sous agonistes de la DA (Bardgett et al., 2009). Ces effets sont de plus indépendants de tout apprentissage et ne sont pas liés à une modification de la valeur subjective de la récompense (Wyvell et Berridge, 2000) : la DA semble ainsi tenir un rôle de modulateur du comportement à l'instar du paramètre de vigueur dans le modèle.

La régulation motivationnelle n'affecte en aucune manière la capacité de contrôle en elle-même. Si une trop faible vigueur entraîne certainement des difficultés pour répondre efficacement à des perturbations extérieures, la coordination motrice est préservée avec toutes ses propriétés fondamentales quel que soit le niveau de vigueur (chapitre 4). En ce sens, le modèle explique pourquoi les manipulations de la DA ou des lésions des ganglions de la base peuvent engendrer une lenteur exceptionnelle des mouvements (bradykinesie) sans pour autant altérer la coordination des actes moteurs (Hikosaka et Wurtz, 1985a,b; Kato et al., 1995; Alamy et al., 1996). Plus particulièrement, le modèle explique pourquoi la destruction des neurones dopaminergiques que provoque la maladie de Parkinson engendre une bradykinésie importante (Marsden, 1989; Robichaud et al., 2002; Negrotti et al., 2005) sans que les autres caractéristiques du mouvement soient nécessairement perturbées (Teasdale et al., 1990; Mazzoni et al., 2007). Cette hypothèse est corroborée par le fait que les traitements médicamenteux qui restaurent l'activité dopaminergique font disparaître la bradykinésie chez les patients parkinsoniens (Negrotti et al., 2005).

6.3.2 Le rôle du décompte

Le comportement, ou tout du moins les actes moteurs, semblent modulables en dehors de toute implication de la DA. Les patients parkinsoniens sont par exemple capables de changer leur vitesse de mouvement si l'instruction leur en est donnée, même s'ils restent globalement plus lents que les sujets sains (Hefter et al., 1996). Un phénomène plus étonnant est celui qui intervient dans les cas d'« urgence motrice ». Dans l'expérience de Majsak et al. (1998) par exemple, les patients parkinsoniens agissent plus vite quand ils ont à stopper à temps une bille qui roule sur une rampe inclinée que quand ils doivent produire un mouvement d'eux-mêmes, sans aucune pression temporelle. En fin de compte, cette capacité à intégrer les contraintes d'urgence dans la spécification de la durée des mouvements s'avère une propriété générale du contrôle moteur (Montagnini et Chelazzi, 2005) qui ne dépend pas de l'intégrité de la transmission dopaminergique (Ballanger et al., 2006) et pourrait ainsi correspondre à l'effet du décompte temporel (γ) dans notre modèle. Il est également possible que la différence de vitesse entre les mouvements auto-initiés et les mouvements déclenchés en réaction à un évènement extérieur (*self-triggered vs. externally-triggered*), phénomène indépendant du circuit dopaminergique (Ballanger et al., 2006), soit également lié au décompte.

Il est toutefois difficile d'identifier avec certitude ces phénomènes non-motivationnels à une contribution spécifique du décompte temporel sur la seule base du comportement moteur, et ces interprétations restent relativement hypothétiques. Le modèle prédit toutefois une relation précise entre mouvement et décision qu'il est possible de tester empiriquement. Cela permettrait de confirmer le rôle du décompte dans les phénomènes que nous avons évoqués. Une expérience possible serait de tester si les conditions d'urgence, en plus d'accélérer les mouvements, biaisent également la décision en faveur des options les moins difficiles comme le prévoit un effet de γ . Autrement dit, notre formalisme prédit que, sous l'influence d'une pression temporelle,

les sujets devraient bouger plus vite mais également avoir tendance à limiter leurs actions aux cibles les plus facilement accessibles même si elles ne sont pas les plus récompensées.

Discussion générale

Afin de survivre, nous devons préserver nos ressources énergétiques avec attention. Chaque action doit être pesée, sélectionnée et exécutée avec soin pour garantir que nos mouvements nous rapportent plus qu'ils ne nous coûtent. Nous avons proposé l'idée qu'à chacun de nos actes nous essayons de maximiser le compromis entre les efforts que nous fournissons et les récompenses que nous obtenons. Ce principe peut être utilisé à tous les niveaux de la préparation du comportement, que ce soit pour décider quel objectif poursuivre ou quel muscle activer. Les résultats présentés montrent que cette hypothèse, aussi simple soit-elle, permet d'expliquer de façon unifiée le mouvement et la décision avec beaucoup de détails. Notre formalisme fournit également un cadre théorique précis pour comprendre les mécanismes en jeu dans la régulation du comportement et plus particulièrement dans la façon dont l'effort est modulé par la motivation. Le modèle donne ainsi des pistes pour expliciter les pathologies qui affectent spécifiquement la capacité du système nerveux central à réguler l'action en fonction du contexte écologique. Ceci présente un enjeu majeur en particulier pour la compréhension du rôle de la dopamine et des ganglions de la base dans l'apparition des symptômes de la maladie de Parkinson.

7.1 Un comportement moteur flexible et unifié

Le modèle proposé s'appuie sur un principe d'optimalité (minimum d'effort) pour expliquer la formation des mouvements volontaires. Notre approche s'inscrit donc

dans la lignée des nombreux travaux précédents qui ont démontré la puissance de la théorie du contrôle optimal dans l'étude de la coordination motrice (Todorov et Jordan, 2002; Todorov, 2004; Scott, 2004; Guigon et al., 2007a, 2008b,a; Diedrichsen et al., 2010). Ce lien de parenté explique le succès de notre modèle dans la reproduction des propriétés du contrôle moteur. Notre formalisme se distingue toutefois de ces approches sur un point fondamental : la gestion du temps. En effet, les modèles actuels postulent que les mouvements sont planifiés pour être exécutés par un contrôleur temporaire sur une durée finie. Dans notre cas le contrôle est permanent, c'est-à-dire que le comportement moteur n'a jamais de fin. La durée de mouvement planifiée qui apparaît dans le calcul du contrôle optimal (équation 2.13) est un artefact de la méthode de résolution que nous utilisons (ce ne serait pas le cas avec un algorithme d'apprentissage par renforcement par exemple) et n'est aucunement une limite réelle de l'exécution motrice. Les durées de mouvement qui sont rapportées dans nos résultats ne sont que des mesures arbitraires de la durée des déplacements du corps et ne reflètent pas une potentielle durée de comportement. Ces propriétés ont plusieurs implications importantes sur la théorie du contrôle moteur qui découle de notre formalisme.

Tout d'abord, l'absence d'une durée prédéterminée permet une complète flexibilité de l'action motrice. Les perturbations de toutes sortes ne sont alors plus des obstacles qui mettent en péril le succès comportemental de l'action, le contrôleur étant toujours prêt à réagir et à guider le mouvement vers sa cible. Ainsi nous n'avons pas besoin de faire appel à des mécanismes extérieurs au contrôle tel qu'un processus de rétro-action mécanique (Shadmehr et Mussa-Ivaldi, 1994) pour expliquer la stabilité du mouvement aux perturbations mécaniques. De façon générale, notre modèle ne fait aucune distinction entre les différents types d'évènements qui peuvent corrompre l'exécution motrice : la même politique de contrôle gouverne uniformément le comportement moteur vers son but, sans interruption, qu'il soit ou non parasité par du

bruit ou des perturbations extérieures.

Ensuite, la politique de contrôle que nous avons modélisée est totalement affranchie du temps qui passe, c'est-à-dire que le choix d'appliquer une commande motrice à un instant donné ne dépend que de l'état actuel du corps (sa position, sa vitesse) et de l'objectif visé. Ceci contraste avec les modèles optimaux classiques (*e.g.* Todorov et Jordan, 2002) où le temps écoulé depuis le début du mouvement détermine très fortement la solution adoptée par le contrôleur. Cette caractéristique permet d'envisager plus facilement comment le contrôleur peut être représentée neuralemement. En effet, comme la politique de contrôle que nous employons ne demande pas une représentation interne du temps pour fonctionner, elle peut être encodée dans un réseau de neurones artificiels relativement simple (Marin et al., 2011). Notre modèle propose ainsi une interprétation nouvelle du codage neuronal du contrôle moteur qu'il serait intéressant de confronter aux données expérimentales encore mal comprises (Scott, 2008; Shadmehr et Krakauer, 2008).

Enfin, nous soulignerons que les modèles classiques du contrôle optimal ont le problème de ne rien dire de ce qui se passe avant et après le mouvement. Ils décrivent en effet les actes moteurs comme des événements parfaitement circonscrits dans le temps, isolés du reste du comportement. Ceci soulève de nombreuses questions quant à la façon dont les mouvements peuvent s'enchaîner, se superposer, et surtout s'intégrer avec un contrôle de la posture supposé indépendant. Ils ne peuvent justifier pourquoi telle contraction musculaire tient du mouvement, telle autre de la posture, et comment le contrôle postural doit reprendre la main quand le contrôle du mouvement se termine à la fin du temps qui lui était imparti. Dans notre formalisme, le comportement moteur dans son ensemble est défini par une unique politique de contrôle optimale qui s'applique à la fois à la gestion des déplacements et à la stabilisation du corps. La séparation entre mouvement et posture n'est alors qu'une distinction factice entre des aspects du comportement moteur qui émergent d'un même processus. Si l'aptitude

de notre modèle à reproduire les propriétés du contrôle postural n'a pas été testée spécifiquement (mais voir Guigon, 2010), sa capacité à unifier posture et mouvement représente une prédiction théorique importante.

7.2 Une redéfinition de la frontière entre décision et mouvement

L'approche computationnelle usuellement utilisée pour expliquer le comportement postule que la décision et la production motrice sont deux processus indépendants l'un de l'autre (Daw et Doya, 2006; Körding, 2007; Rangel et al., 2008), le premier visant à maximiser les récompenses obtenues et le second à minimiser le coût de réalisation de l'action. Cette approche a deux défauts majeurs. Tout d'abord, les modèles de décision ne peuvent rendre compte précisément de l'effet du coût d'exécution des mouvements sur le choix des actions. Réciproquement, les modèles du contrôle moteur offrent une description du mouvement encore difficile à intégrer dans une vision écologique de l'action et peinent à expliquer les influences du contexte motivationnel sur le mouvement. Pour remédier à ces problèmes, nous avons proposé l'idée que la décision et le contrôle moteur sont intimement liés par un objectif unique où récompenses et efforts sont conjointement optimisés. Cette hypothèse permet tout d'abord de capturer l'influence très générale de l'effort sur la décision (Stephens et Krebs, 1986; Stevens et al., 2005; Walton et al., 2006; Croxson et al., 2009). Puisque le coût de l'action est anticipé à partir d'une planification précise du mouvement, ce modèle garantit la justesse de l'évaluation de l'effort à venir et permet ainsi d'expliquer la décision dans des situations où le coût de l'action doit être mesuré avec une grande précision (Cos et al., 2011). Ensuite, notre formalisme permet de rendre compte naturellement de la perméabilité du mouvement au contexte motivationnel. Plus précisément, il donne une raison explicite à l'effet énergisant des récompenses

sur les actes moteurs en déterminant à la volée l'effort à investir dans une action en fonction de la valeur de l'objectif visé. Cette allocation dynamique de l'effort pour répondre à une stimulation motivationnelle affecte à la fois le mouvement, en jouant sur la vitesse d'exécution de l'action, et la décision, en régulant l'ardeur qui peut être accordée à la poursuite d'un objectif.

De façon plus générale, notre modèle fait tomber la barrière entre comportement moteur et décision. Ceci permet d'aborder de façon normative les tâches qui ne peuvent être clairement catégorisées et sur lesquelles les approches formelles courantes sont vite limitées. Nous citerons par exemple la production de mouvements contraints par des points de passage intermédiaires, à la fois séquence comportementale et acte moteur élaboré.

Le modèle a été appliqué à des tâches purement motrices qui n'impliquent aucune récompense explicite (Shadmehr et Mussa-Ivaldi, 1994; Liu et Todorov, 2007). Bien que ces tâches ne correspondent pas à première vue à des situations de compromis coût/bénéfice, des études expérimentales montrent cependant que leur exécution est clairement modulée par des facteurs liés aux coûts et aux bénéfices, *e.g.* les charges (Bock, 1990), la fatigue (Corcos et al., 2002), la difficulté de la tâche (Montagnini et Chelazzi, 2005), ou l'attractivité des stimuli (Xu-Wilson et al., 2009). Ces observations suggèrent que les tâches purement motrices pourraient partager les mêmes représentations sous-jacentes que les tâches impliquant des récompenses explicites. Plus généralement, nous pouvons nous interroger sur l'existence d'une motivation implicite qui sous-tendrait l'ensemble de notre comportement (Veltkamp et al., 2009).

Une proposition centrale du modèle est l'existence d'une base commune pour la décision et l'action. Les seules données expérimentales qui apportent un réel soutien au modèle sont celles de Stevens et al. (2005) et de Cos et al. (2011). L'élément-clé est l'observation combinée des caractéristiques de choix et des actions motrices associées. En fait, toute tâche impliquant une situation de compromis coût/bénéfice pourrait

être utilisée si la durée des déplacements est enregistrée (ce qui n'est pas le cas dans les études publiées). Dans ce cadre, la méthodologie développée par Cos et al. (2011) pourrait être étendue à des tâches impliquant des récompenses.

7.2.1 Conséquences pour l'architecture du contrôle du comportement

La littérature conçoit souvent la décision et l'exécution motrice comme des processus sériels. Autrement dit, elle postule que le cerveau décide *où* aller avant de définir *comment* y aller. La fusion des objectifs de la décision et de la motricité remet partiellement en question cette hiérarchie. Selon notre hypothèse en effet, les actes moteurs sont nécessairement planifiés avant que toute décision ne soit possible. Si plusieurs objectifs sont envisageables, les diverses actions possibles doivent toutes être représentées avec précision en amont du processus de sélection. Ce dernier, après évaluation des bénéfices mais aussi des coûts de chaque option, a pour tâche de conserver le meilleur plan moteur et de lancer son exécution. Si cette idée est compatible avec un nombre grandissant de données expérimentales (Cisek, 2007), elle se distingue de la théorie du renforcement qui sert traditionnellement de cadre formel pour expliquer les fonctions cérébrales associées au contrôle du comportement.

Le rôle de la dopamine

Selon la théorie du renforcement, la dopamine guide la décision permettant d'évaluer chaque action à hauteur de la récompense qu'elle vise. L'hypothèse sous-jacente à ce principe est que l'utilité d'une action est égale à la récompense qu'elle permet d'obtenir. Si l'activité dopaminergique corrèle effectivement avec la valeur anticipée d'une récompense (Fiorillo et al., 2003; Schultz et al., 2008; Schultz, 2010), l'idée qu'elle code également la valeur des actions présente quelques incohérences. Tout d'abord, si la DA code véritablement l'utilité des actes moteurs nous devrions observer une

modulation de son activité avec la quantité d'effort à venir. Cette hypothèse a toutefois été invalidée par plusieurs travaux récents (Wanat et al., 2010; Day et al., 2010; Gan et al., 2010). Plus problématique, cette vision de la DA ne peut expliquer ses effets sur le contenu des actions (la durée des mouvements par exemple) ni pourquoi ce neurotransmetteur est si important dans les situations d'effort.

Dans notre modèle, nous avons identifié la DA à un *paramètre* de l'action, la vigueur. Cette hypothèse permet de capturer l'effet tonifiant de la DA sur le comportement. Plus précisément, notre formalisme explique l'importance de la DA pour surmonter les efforts requis (Walton et al., 2006; Phillips et al., 2007; Floresco et al., 2008; St Onge et Floresco, 2009; Croxson et al., 2009; Gan et al., 2010). Le modèle donne ainsi une interprétation formelle de la réticence à produire des efforts moteurs que produit une déplétion de la DA, une théorie particulièrement intéressante pour comprendre par exemple la bradykinésie qu'occasionne la maladie de Parkinson (Mazzoni et al., 2007). Plus généralement, cette théorie place ainsi la DA comme le levier capable d'adapter le comportement en fonction de l'intérêt écologique de la tâche par rapport à l'état interne de l'animal. En relativisant la valeur des ressources disponibles en fonction de l'information fournie par l'environnement, de l'expérience acquise, ou des besoins immédiats, le rôle de la DA pourrait être ainsi de moduler l'effort à investir dans l'action en fonction du contexte dans lequel s'exprime le comportement (Ostlund et al., 2011).

Un débat existe actuellement pour savoir si la DA affecte préférentiellement la pondération des efforts (Mazzoni et al., 2007) ou l'évaluation des récompenses (Nicola, 2010). Si notre modèle ne peut distinguer entre ces deux hypothèses (à cause de la symétrie entre ρ et η), il propose toutefois une explication à cette ambivalence de la DA en explicitant l'interaction réciproque entre efforts et récompenses.

La fonction des ganglions de la base

L'information véhiculée par la dopamine est majoritairement traitée par les ganglions de la base (GB). Ces derniers ont ainsi un rôle primordial dans le contrôle motivationnel du comportement, comme l'a largement souligné la littérature. Dans le cadre de notre théorie, les GB pourraient intégrer les informations motivationnelles véhiculées par la dopamine et se charger d'évaluer dynamiquement le compromis entre coûts moteurs et bénéfices escomptés. Les noyaux d'entrée des GB semblent en effet construire une représentation globale des actions possibles dans leur contexte environnemental (Pasquereau et al., 2007; Garenne et al., 2011), c'est-à-dire d'agréger les informations concernant les récompenses (Samejima et al., 2005; Opris et al., 2011) et les efforts (Croxson et al., 2009; Kurniawan et al., 2010) associées aux actions possibles dans la tâche en cours. Les computations qui se déroulent au sein des GB permettraient alors de produire au niveau de leur sortie (au niveau du globus pallidus interne) un signal à même d'ajuster un plan moteur représenté dans le cortex, *i.e.* de paramétrer le contrôleur de l'action, de le tempérer ou de le stimuler dynamiquement en fonction de la situation écologique affrontée (Turner et al., 1998; Pessiglione et al., 2007; Turner et Desmurget, 2010; Grafton et Tunik, 2011). L'idée que les GB participent ainsi activement à la régulation de la vitesse des mouvements (de leur effort) en fonction de la motivation a été particulièrement bien mise en évidence dans le cas des saccades oculaires (Nakamura et Hikosaka, 2006; Hikosaka, 2007). La spécificité de ce processus apparaît clairement dans le cas de lésions des GB qui produisent une incapacité à réguler l'effort en fonction de la récompense sans toutefois perturber ni l'aptitude du patient à répondre aux exigences de la tâche (Schmidt et al., 2008, voir figure 7.1) ni sa faculté de coordination motrice (Desmurget et Turner, 2008, 2010).

La question de l'implémentation computationnelle de la régulation du mouvement par les GB reste toutefois ouverte. L'hypothèse que les GB établissent un compromis entre efforts et bénéfices n'a fait l'objet que de très rares travaux de modélisa-

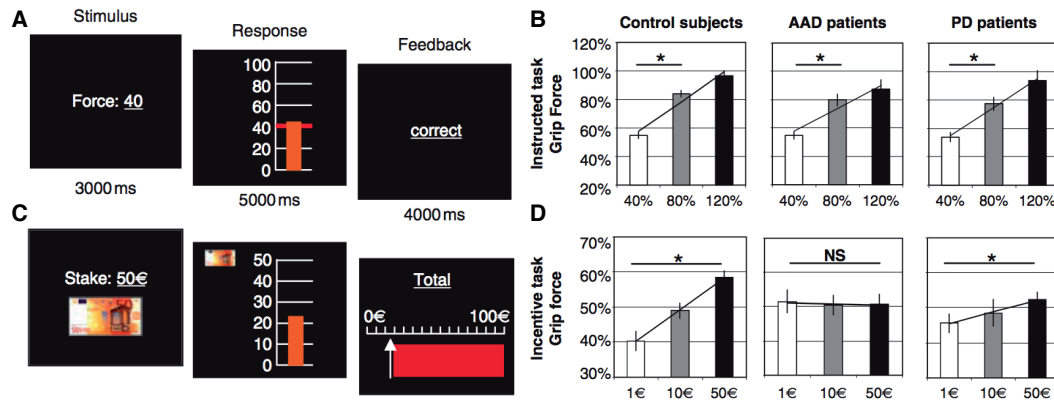


FIGURE 7.1 – *Expérience de Schmidt et al. (2008)*. Dans une première condition (A, *instructed task*), des sujets doivent presser une poignée afin de faire monter à un niveau précis une jauge affichée sur un écran, ce qu'ils arrivent toujours à faire (B) en dépit de leur pathologie éventuelle (AAD : lésion bilatérale des GB, PD : maladie de Parkinson). Dans une autre condition (C, *incentive task*), les sujets peuvent gagner une mise (1, 10 ou 50€) à hauteur de la force qu'ils exercent. Alors que les sujets sains régulent leur effort en fonction de la mise, les patients parkinsoniens sont plus modérés. Les patients lésés, eux, ne modulent absolument pas leur comportement en fonction de la mise (D).

tion (Parush et al., 2011). Cette idée pourrait pourtant être une piste intéressante pour approfondir ou dépasser la théorie plus populaire qui prête aux GB le rôle de décider entre des actions discrètes (Joel et al., 2002; Daw et Doya, 2006; Balleine et O'Doherty, 2010). Plus généralement, nous pouvons questionner la nature binaire souvent admise du signal émis les GB, qui semble plutôt « colorer », en fonction du contexte comportemental de l'action, les plans moteurs représentés au niveau cortical (Kao et al., 2005; Schmidt et al., 2008; Leblois et al., 2010; Turner et Desmurget, 2010).

7.2.2 Le paramètre de décompte

Les effets du décompte temporel sont beaucoup plus délicats à cerner que les effets motivationnels. Ceci peut paraître surprenant au regard de l'abondance de la littérature utilisant la notion de décompte, en particulier en psychologie et en économie. Toutefois, il semble que la notion de décompte utilisée dans ces études ne capture pas

exactement le mécanisme de dévaluation à l'œuvre dans la régulation de la durée des mouvements. En effet, alors que dans notre cas le décompte agit sur la durée de l'action (sur laquelle un contrôle est possible), le concept de décompte est habituellement utilisé pour caractériser la dévaluation qu'occasionne une attente passive de la récompense, après la fin de l'action. L'expérience de Stevens et al. (2005) permet d'éclaircir cette différence. En plus de la tâche que nous avons décrite dans le chapitre 5, les auteurs ont en effet étudié le comportement des singes dans une tâche de choix entre une petite récompense immédiate et une grande récompense donnée après un certain délai, une tâche classique d'attente passive. De façon assez surprenante, le degré de patience que les différents singes montrent lors de l'attente passive des récompenses ne prédit aucunement leur comportement moteur et décisionnel dans la tâche motrice. Les marmousets par exemple s'avèrent très patients alors qu'il préfèrent les récompenses les plus proches (et donc les plus rapides à obtenir). Si cette différence peut éventuellement être attribuée à un effet du coût du déplacement dans la tâche motrice, ce résultat souligne la difficulté d'isoler la contribution du décompte dans l'élaboration du comportement. De façon générale, le phénomène de dévaluation temporelle semble provenir d'un enchevêtrement de processus encore mal compris (Berns et al., 2007), et le rôle spécifique d'une certaine forme de décompte dans le contrôle du comportement moteur reste à élucider.

Une autre remarque concerne la forme du décompte. Il existe en effet un grand débat pour savoir si les récompenses sont décomptées avec le temps de façon exponentielle (comme dans notre modèle) ou hyperbolique (Green et Myerson, 1996). Le décompte exponentiel est nécessaire à la formulation mathématique du modèle car il permet d'obtenir une formulation récursive et invariante dans le temps (Doya, 2000; Todorov, 2007). Cependant, il ne permet pas de rendre compte des incohérences dans les choix inter-temporels qui sont par contre très bien expliquées par le décompte hyperbolique (Green et Myerson, 1996; Sozou, 1998). Les implications

pour le modèle ne sont évidentes à comprendre. Nous pouvons cependant faire deux remarques. Premièrement, nous avons testé notre modèle avec un décompte hyperbolique (au détriment de la rigueur mathématique; voir cependant Alexander et Brown, 2010) et nous n'avons pas trouvé de différences visibles sur les propriétés du modèle. Deuxièmement, comme montré ci-dessus à partir des résultats de Stevens et al. (2005), le décompte dans le modèle semble être différent du décompte mesuré dans des tâches inter-temporelles (*delay discounting*) et pourrait être spécifique des situations de compromis coût/bénéfice et du contrôle moteur. En particulier, il existe de nombreuses observations expérimentales montrant que la prise de décision dans le cadre du contrôle moteur est beaucoup plus normative (*i.e.* beaucoup moins sujette à des déviations par rapport à l'optimalité) que la prise de décision en économie (Trommershäuser et al., 2008). Dans ce sens, un décompte exponentiel pourrait se justifier dans le modèle.

7.3 Conclusion

Ce travail a proposé une approche computationnelle unifiée de la décision et de l'action. Elle repose sur une formalisation en terme de choix optimal pour des situations de type coût/bénéfice et exploite l'idée d'une cohérence nécessaire entre décision et action. Autrement dit, si un sujet prend la décision d'exécuter une action qui lui semble la meilleure en terme de bénéfices nets, il faut que les coûts liés à l'exécution soient conformes à ce qui avait été anticipé. Ce principe de cohérence mène à une imbrication complète des processus de décision et d'action. La décision et le mouvement deviennent alors indissociables, et le choix entre des options distinctes autant que la production continue de l'action proviennent d'une même optimisation permanente du meilleur compromis entre efforts et bénéfices.

Pour valider ce modèle, nous avons étudié son comportement dans des tâches pu-

rement motrices et dans des situations typiques de compromis coût/bénéfice. Nous avons ainsi montré que le modèle produisait une explication normative de ces comportements. Une force du modèle est d'expliquer précisément la façon dont l'effort est ajusté en fonction de la situation écologique du comportement. Sur cette base, nous avons tissé des liens entre les paramètres du modèle et l'influence de la dopamine sur les processus de décision et d'action. Une proposition importante est une association/dissociation possible entre facteur motivationnel et processus dopaminergique d'un côté, et entre facteur de décompte temporel et processus non-dopaminergique de l'autre. Cette dissociation, dont une dimension bien connue pourrait correspondre aux kinèses paradoxales, est une proposition nouvelle de ce travail qui ouvre des perspectives intéressantes dans l'étude des fonctions normales et pathologiques de la dopamine et des ganglions de la base.

Bibliographie

- Aarts, H., Custers, R., et Marien, H. (2008). Preparing and motivating behavior outside of awareness. *Science*, 319(5870), 1639.
- Adam, J. J., Nieuwenstein, J. H., Huys, R., Paas, F. G. W. C., Kingma, H., Willems, P., et Werry, M. (2000). Control of rapid aimed hand movements : The one-target advantage. *Journal of Experimental Psychology : Human Perception and Performance*, 26(1), 295–312.
- Alamy, M., Pons, J., Gambarelli, D., et Trouche, E. (1996). A defective control of small amplitude movements in monkeys with globus pallidus lesions : An experimental study on one component of pallidal bradykinesia. *Behavioral Brain Research*, 72(1-2), 57–62.
- Alexander, W. H. et Brown, J. W. (2010). Hyperbolically discounted temporal difference learning. *Neural Computation*, 22(6), 1511–1527.
- Bahill, A., Clark, M., et Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3-4), 191–204.
- Bainbridge, L. et Sanders, M. (1972). The generality of Fitts's law. *Journal of Experimental Psychology*, 96(1), 130–133.
- Bakker, M., Lange, F. P., Stevens, J. A., Toni, I., et Bloem, B. R. (2007). Motor imagery of gait : A quantitative approach. *Experimental Brain Research*, 179(3), 497–504.
- Ballanger, B., Thobois, S., Baraduc, P., Turner, R. S., Broussolle, E., et Desmurget, M. (2006). "Paradoxical kinesis" is not a hallmark of Parkinson's disease but a general property of the motor system. *Movement Disorders*, 21(9), 1490–1495.
- Balleine, B. W. et O'Doherty, J. P. (2010). Human and rodent homologies in action control : Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35(1), 48–69.
- Bardgett, M. E., Depenbrock, M., Downs, N., Points, M., et Green, L. (2009). Dopamine modulates effort-based decision making in rats. *Behavioral Neuroscience*, 123(2), 242–251.

- Beamish, D., MacKenzie, I. S., et Wu, J. (2006). Speed-accuracy trade-off in planned arm movements with delayed feedback. *Neural Networks*, 19(5), 582–599.
- Berns, G. S., Laibson, D. I., et Loewenstein, G. (2007). Intertemporal choice—toward an integrative framework. *Trends in Cognitive Sciences*, 11(11), 482–488.
- Bernstein, N. (1967). *The Coordination and Regulation of Movements*. Oxford, UK : Pergamon Press.
- Berridge, K. C. et Aldridge, J. W. (2008). Decision utility, the brain, and pursuit of hedonic goals. *Social Cognition*, 26(5), 621–646.
- Berridge, K. C. et Aldridge, J. W. (2009). Decision utility, incentive salience, and cue-triggered “wanting”. In E. Morsella, J. Bargh, et P. Gollwitzer (Eds.), *Oxford Handbook of Human Action* (pp. 509–533). New York, NY : Oxford University Press.
- Berridge, K. C. et Robinson, S. (1998). What is the role of dopamine in reward : Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28(3), 309–369.
- Berridge, K. C. et Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, 26(9), 507–513.
- Berridge, K. C., Robinson, T. E., et Aldridge, J. W. (2009). Dissecting components of reward : ‘liking’, ‘wanting’, and learning. *Current Opinion in Pharmacology*, 9(1), 65–73.
- Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control*. Belmont, MA : Athena Scientific.
- Bhushan, N. et Shadmehr, R. (1999). Computational nature of human adaptive control during learning of reaching movements in force fields. *Biological Cybernetics*, 81(1), 39–60.
- Bock, O. (1990). Load compensation in human goal-directed arm movements. *Behavioural Brain Research*, 41(3), 167–177.
- Brown, J. et Slater-Hammel, A. (1949). Discrete movements in the horizontal plane as a function of their length and direction. *Journal of Experimental Psychology*, 39(1), 84–95.

- Brown, S. H., Hefter, H., Mertens, M., et Freund, H. J. (1990). Disturbances in human arm movement trajectory due to mild cerebellar dysfunction. *Journal of Neurology, Neurosurgery, and Psychiatry*, 53(4), 306–313.
- Bryson, A. E. et Ho, Y.-C. (1975). *Applied Optimal Control - Optimization, Estimation, and Control*. New York, NY : Hemisphere.
- Burdet, E., Osu, R., Franklin, D. W., Milner, T. E., et Kawato, M. (2001). The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, 414(6862), 446–449.
- Burdet, E., Tee, K., Mareels, I., Milner, T. E., Chew, C., Franklin, D., Osu, R., et Kawato, M. (2006). Stability and motor adaptation in human arm movements. *Biological Cybernetics*, 94(1), 20–32.
- Cagniard, B., Balsam, P. D., Brunner, D., et Zhuang, X. (2006). Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. *Neuropsychopharmacology*, 31(7), 1362–1370.
- Cisek, P. (2007). Cortical mechanisms of action selection : The affordance competition hypothesis. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 362(1485), 1585–1599.
- Cooke, J. (1980). The organization of simple, skilled movements. In G. E. Stelmach et L. Requin (Eds.), *Tutorials in Motor Behavior, Advances in Psychology* (pp. 199–212). Amsterdam : North-Holland.
- Corcos, D. M., Jiang, H.-Y., Wilding, J., et Gottlieb, G. L. (2002). Fatigue induced changes in phasic muscle activation patterns for fast elbow flexion movements. *Experimental Brain Research*, 142(1), 1–12.
- Cos, I., Belanger, N., et Cisek, P. (2011). The influence of predicted arm biomechanics on decision making. *Journal of Neurophysiology*, 105(6), 3022–3033.
- Cousins, M. S. et Salamone, J. D. (1994). Nucleus accumbens dopamine depletions in rats affect relative response allocation in a novel cost/benefit procedure. *Pharmacology, Biochemistry, and Behavior*, 49(1), 85–91.
- Crespi, L. (1942). Quantitative variation in incentive and performance in the white rat. *The American Journal of Psychology*, 55(4), 467–517.

- Crosson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J., et Rushworth, M. F. S. (2009). Effort-based cost-benefit valuation and the human brain. *The Journal of Neuroscience*, 29(14), 4531–4541.
- Daw, N. D. et Doya (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, 16(2), 199–204.
- Daw, N. D. et Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Social Cognition*, 26(5), 593–620.
- Day, J. J., Jones, J. L., Wightman, R. M., et Carelli, R. M. (2010). Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biological Psychiatry*, 68(3), 306–309.
- Dayan, P., Niv, Y., Seymour, B., et Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Networks*, 19(8), 1153–1160.
- Denk, F., Walton, M. E., Jennings, W., Sharp, T., Rushworth, M. F. S., et Bannerman, D. M. (2005). Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology*, 179(3), 587–596.
- Desmurget, M. et Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11), 423–431.
- Desmurget, M. et Turner, R. S. (2008). Testing basal ganglia motor functions through reversible inactivations in the posterior internal globus pallidus. *Journal of Neurophysiology*, 99(3), 1057–1076.
- Desmurget, M. et Turner, R. S. (2010). Motor sequences and the basal ganglia : Kinematics, not habits. *The Journal of Neuroscience*, 30(22), 7685–7690.
- Diedrichsen, J., Shadmehr, R., et Ivry, R. B. (2010). The coordination of movement : Optimal feedback control and beyond. *Trends in Cognitive Sciences*, 14(1), 31–39.
- Domjan, M. (2002). *The Principles of Learning and Behavior*. Belmont, CA : Wadsworth Publishing.
- Dorfman, R. (1969). An economic interpretation of optimal control theory. *The American Economic Review*, 59(5), 817–31.
- Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural Computation*, 12(1), 219–245.

- Engelbrecht (2001). Minimum principles in motor control. *Journal of Mathematical Psychology*, 45(3), 497–454.
- Fiorillo, C. D., Tobler, P. N., et Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614), 1898–1902.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.
- Flanagan, J. R., Vetter, P., Johansson, R., et Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology*, 13(2), 146–150.
- Flash, T. et Hogan, N. (1985). The coordination of arm movements : An experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5(7), 1688–1703.
- Floresco, S. B., St Onge, J. R., Ghods-Sharifi, S., et Winstanley, C. A. (2008). Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 375–389.
- Frederick, S., Loewenstein, G., et O’Donoghue, T. (2002). Time discounting and time preference : A critical review. *Journal of Economic Literature*, 40(2), 351–401.
- Gan, J. O., Walton, M. E., et Phillips, P. E. M. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature Neuroscience*, 13(1), 25–27.
- Garenne, A., Pasquereau, B., Guthrie, M., Bioulac, B., et Boraud, T. (2011). Basal ganglia preferentially encode context dependent choice in a two-armed bandit task. *Frontiers in Systems Neuroscience*, 5, 23.
- Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning : The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108 Suppl 3, 15647–15654.
- Glimcher, P. W., Dorris, M., et Bayer, H. (2005). Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior*, 52(2), 213–256.
- Glimcher, P. W., Kable, J. W., et Louie, K. (2007). Neuroeconomic studies of impulsivity : Now or just as soon as possible? *The American Economic Review*, 97(2), 142–147.

- Goodale, M. A., Pélisson, D., et Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320(6064), 748–750.
- Gordon, J. et Ghez, C. (1984). EMG patterns in antagonist muscles during isometric contraction in man - Relations to response dynamics. *Experimental Brain Research*, 55(1), 167–171.
- Gordon, J., Ghilardi, M. F., Cooper, S. E., et Ghez, C. (1994a). Accuracy of planar reaching movements. II. Systematic extent errors resulting from inertial anisotropy. *Experimental Brain Research*, 99(1), 112–130.
- Gordon, J., Ghilardi, M. F., et Ghez, C. (1994b). Accuracy of planar reaching movements. I. Independence of direction and extent variability. *Experimental Brain Research*, 99(1), 97–111.
- Grafton, S. et Tunik, E. (2011). Human basal ganglia and the dynamic control of force during on-line corrections. *Journal of Neuroscience*, 31(5), 1600–1605.
- Green, L. et Myerson, J. (1996). Exponential versus hyperbolic discounting of delayed outcomes : Risk and waiting times. *American Zoology*, 36(4), 496–505.
- Gribble, P. L., Mullin, L. I., Cothros, N., et Mattar, A. (2003). Role of cocontraction in arm movement accuracy. *Journal of Neurophysiology*, 89(5), 2396–2405.
- Guigon, E. (2010). Active control of bias for the control of posture and movement. *Journal of Neurophysiology*, 104(2), 1090–1102.
- Guigon, E. (2011). Models and architectures for motor control : Simple or complex ? In F. Danion et M. Latash (Eds.), *Motor Control* (pp. 478–502). Oxford, UK : Oxford University Press.
- Guigon, E., Baraduc, P., et Desmurget, M. (2007a). Computational motor control : Redundancy and invariance. *Journal of Neurophysiology*, 97(1), 331–347.
- Guigon, E., Baraduc, P., et Desmurget, M. (2007b). Coding of movement- and force-related information in primate primary motor cortex : A computational approach. *The European Journal of Neuroscience*, 26(1), 250–260.
- Guigon, E., Baraduc, P., et Desmurget, M. (2008a). Optimality, stochasticity, and variability in motor behavior. *Journal of Computational Neuroscience*, 24(1), 57–68.

- Guigon, E., Baraduc, P., et Desmurget, M. (2008b). Computational motor control : Feedback and accuracy. *The European Journal of Neuroscience*, 27(4), 1003–1016.
- Hallett, M. et Khoshbin, S. (1980). A physiological mechanism of bradykinesia. *Brain : a journal of neurology*, 103(2), 301–314.
- Harris, C. M. et Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695), 780–784.
- Harris, C. M. et Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological Cybernetics*, 95(1), 21–29.
- Hefter, H., Brown, S. H., Cooke, J. D., et Freund, H. J. (1996). Basal ganglia and cerebellar impairment differentially affect the amplitude and time scaling during the performance of forearm step tracking movements. *Electromyography and clinical neurophysiology*, 36(2), 121–128.
- Hicheur, H., Pham, Q.-C., Arechavaleta, G., Laumond, J.-P., et Berthoz, A. (2007). The formation of trajectories during goal-oriented locomotion in humans. I. A stereotyped behaviour. *The European Journal of Neuroscience*, 26(8), 2376–2390.
- Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. *Annals of the New York Academy of Sciences*, 1104, 229–249.
- Hikosaka, O. et Wurtz, R. H. (1985a). Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus. *Journal of Neurophysiology*, 53(1), 266–291.
- Hikosaka, O. et Wurtz, R. H. (1985b). Modification of saccadic eye movements by GABA-related substances. II. Effect of muscimol in monkey substantia nigra pars reticulata. *Journal of Neurophysiology*, 53(1), 292–308.
- Hoff, B. (1994). A model of duration in normal and perturbed reaching movement. *Biological Cybernetics*, 71(6), 481–488.
- Izawa, J., Rane, T. D., Donchin, O., et Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *The Journal of Neuroscience*, 28(11), 2883–2891.
- Jeannerod, M. (1988). *The Neural and Behavioral Organization of Goal-Directed Movements*. Oxford, UK : Clarendon Press.

- Joel, D., Niv, Y., et Ruppin, E. (2002). Actor-critic models of the basal ganglia : New anatomical and computational perspectives. *Neural Networks*, 15(4-6), 535–547.
- Jordan, M. (1996). Computational aspects of motor control and motor learning. In H. Heuer et S. Keele (Eds.), *Handbook of Perception and Action, Vol 2 : Motor Skill* (pp. 71–118). New York, NY : Academic Press.
- Kandel, E. R., Schwartz, J., et Jessel, T. (2000). *Principles of Neural Science*. New York, NY : McGraw-Hill, 4^o edition.
- Kao, M. H., Doupe, A., et Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, 433(7026), 638–643.
- Kato, M., Miyashita, N., Hikosaka, O., Matsumura, M., Usui, S., et Kori, A. (1995). Eye movements in monkeys with local dopamine depletion in the caudate nucleus. I. Deficits in spontaneous saccades. *The Journal of Neuroscience*, 15(1 Pt 2), 912–927.
- Khan, M., Lawrence, G., Franks, I., et Buckolz, E. (2004). The utilization of visual feedback from peripheral and central vision in the control of direction. *Experimental Brain Research*, 158(2), 241–251.
- Kirk, D. E. (2004). *Optimal Control Theory : An Introduction*. Mineola, NY : Dover.
- Kobayashi, S., Nomoto, K., Watanabe, M., Hikosaka, O., Schultz, W., et Sakagami, M. (2006). Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. *Neuron*, 51(6), 861–870.
- Körding, K. P. (2007). Decision theory : What “should” the nervous system do? *Science*, 318, 606–610.
- Körding, K. P. et Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7), 319–326.
- Krakauer, J. W. et Shadmehr, R. (2007). Towards a computational neuropsychology of action. *Progress in Brain Research*, 165, 383–394.
- Kunz, B. R., Creem-Regehr, S. H., et Thompson, W. B. (2009). Evidence for motor simulation in imagined locomotion. *Journal of Experimental Psychology : Human Perception and Performance*, 35(5), 1458–1471.

- Kurniawan, I., Seymour, B., Talmi, D., Yoshida, W., Chater, N., et Dolan, R. J. (2010). Choosing to make an effort : The role of striatum in signaling physical effort of a chosen action. *Journal of Neurophysiology*, 104(1), 313–321.
- Kurniawan, I. T., Guitart-Masip, M., et Dolan, R. J. (2011). Dopamine and effort-based decision making. *Frontiers in Neuroscience*, 5, 81.
- Leblois, A., Wendel, B. J., et Perkel, D. J. (2010). Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *The Journal of Neuroscience*, 30(16), 5730–5743.
- Liu, D. et Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *The Journal of Neuroscience*, 27(35), 9354–9368.
- Majsak, M. J., Kaminski, T., Gentile, A. M., et Flanagan, J. R. (1998). The reaching movements of patients with Parkinson’s disease under self-determined maximal speed and visually cued conditions. *Brain : a journal of neurology*, 121 (Pt 4), 755–766.
- Marin, D., Decock, J., Rigoux, L., et Sigaud, O. (2011). Learning cost-efficient control policies with XCSF : Generalization capabilities and further improvement. In *GECCO* (pp. 1–8).
- Marsden, A. M. (1989). Slowness of movement in Parkinson’s disease. *Movement Disorders*, 4 Suppl 1, S26–37.
- Mazzoni, P., Hristova, A., et Krakauer, J. W. (2007). Why don’t we move faster? Parkinson’s disease, movement vigor, and implicit motivation. *The Journal of Neuroscience*, 27(27), 7105–7116.
- McCloskey, D. I. (1978). Kinesthetic sensibility. *Physiological Reviews*, 58(4), 763–820.
- McClure, S. M., Daw, N. D., et Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, 26(8), 423–428.
- Messier, J. et Kalaska, J. F. (1999). Comparison of variability of initial kinematics and endpoints of reaching movements. *Experimental Brain Research*, 125(2), 139–152.
- Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., et Smith, J. E. (1988). Optimality in human motor performance : Ideal control of rapid aimed movements. *Psychological Review*, 95(3), 340–370.

- Montagnini, A. et Chelazzi, L. (2005). The urgency to look : Prompt saccades to the benefit of perception. *Vision Research*, 45(27), 3391–3401.
- Montague, P. R. et Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36(2), 265–284.
- Morishige, K., Osu, R., Miyamoto, H., et Kawato, M. (2006). The sources of variability in the time course of reaching movements. In *2nd International Conference on Brain-inspired Information Technology, International Congress Series* (pp. 105–108).
- Mühlhoff, N., Stevens, J. R., et Reader, S. M. (2011). Spatial discounting of food and social rewards in guppies (*Poecilia reticulata*). *Frontiers in Psychology*, 2, 68.
- Nagasaki, H. (1989). Asymmetric velocity and acceleration profiles of human arm movements. *Experimental Brain Research*, 74(2), 319–326.
- Nakamura, K. et Hikosaka, O. (2006). Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *The Journal of Neuroscience*, 26(20), 5360–5369.
- Negrotti, A., Secchi, C., et Gentilucci, M. (2005). Effects of disease progression and L-dopa therapy on the control of reaching-grasping in Parkinson’s disease. *Neuropsychologia*, 43(3), 450–459.
- Nicola, S. M. (2010). The flexible approach hypothesis : Unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. *Journal of Neuroscience*, 30(49), 16585–16600.
- Nijhof, E.-J. (2003). On-line trajectory modifications of planar, goal-directed arm movements. *Human Movement Science*, 22(1), 13–36.
- Niv, Y., Joel, D., et Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Sciences*, 10(8), 375–381.
- Opris, I., Lebedev, M., et Nelson, R. J. (2011). Motor planning under unpredictable reward : Modulations of movement vigor and primate striatum activity. *Frontiers in Decision Neuroscience*, 5, 61.

- Ostlund, S. B., Wassum, K. M., Murphy, N. P., Balleine, B. W., et Maidment, N. T. (2011). Extracellular dopamine levels in striatal subregions track shifts in motivation and response cost during instrumental conditioning. *Journal of Neuroscience*, 31(1), 200–207.
- Osu, R., Kamimura, N., Iwasaki, H., Nakano, E., Harris, C. M., Wada, Y., et Kawato, M. (2004). Optimal impedance control for task achievement in the presence of signal-dependent noise. *Journal of Neurophysiology*, 92(2), 1199.
- Osu, R., Morishige, K.-i., Miyamoto, H., et Kawato, M. (2009). Feedforward impedance control efficiently reduce motor variability. *Neuroscience Research*, 65(1), 6–10.
- Parush, N., Tishby, N., et Bergman, H. (2011). Dopaminergic balance between reward maximization and policy complexity. *Frontiers in Systems Neuroscience*, 5, 22.
- Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C. E., et Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. *The Journal of Neuroscience*, 27(5), 1176–1183.
- Peciña, S., Cagniard, B., Berridge, K. C., Aldridge, J. W., et Zhuang, X. (2003). Hyperdopaminergic mutant mice have higher "wanting" but not "liking" for sweet rewards. *The Journal of Neuroscience*, 23(28), 9395–9402.
- Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., et Frith, C. D. (2007). How the brain translates money into force : A neuroimaging study of subliminal motivation. *Science*, 316(5826), 904–906.
- Phillips, P. E. M., Walton, M. E., et Jhou, T. C. (2007). Calculating utility : Preclinical evidence for cost-benefit analysis by mesolimbic dopamine. *Psychopharmacology*, 191(3), 483–495.
- Platt, M. L. et Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741), 233–238.
- Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., et Mishchenko, E. F. (1986). *L. S. Pontryagin Selected Works, Volume 4 : The Mathematical Theory of Optimal Processes*. New York, NY : Gordon and Breach.
- Prablanc, C., Desmurget, M., et Gréa, H. (2003). Neural control of on-line guidance of hand reaching movements. *Progress in Brain Research*, 142, 155–170.

- Prablanc, C., Pélisson, D., et Goodale, M. A. (1986). Visual control of reaching movements without vision of the limb. I. Role of retinal feedback of target position in guiding the hand. *Experimental Brain Research*, 62(2), 293–302.
- Press, W., Teukolsky, S., et Vetterling, W. (1992). *Numerical recipes in C. The Art of Scientific Computing*. New York, NY : Cambridge University Press, 2^o édition.
- Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., et Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *The Journal of Neuroscience*, 30(42), 14080.
- Rangel, A., Camerer, C. F., et Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556.
- Robichaud, J. A., Pfann, K. D., Comella, C. L., et Corcos, D. M. (2002). Effect of medication on EMG patterns in individuals with Parkinson’s disease. *Movement Disorders*, 17(5), 950–960.
- Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., et Rushworth, M. F. S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience*, 9(9), 1161–1168.
- Salamone, J. D. et Correa, M. (2002). Motivational views of reinforcement : Implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behavioural Brain Research*, 137(1-2), 3–25.
- Salamone, J. D., Correa, M., Farrar, A., et Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology*, 191(3), 461–482.
- Salamone, J. D., Correa, M., Farrar, A., Nunes, E. J., et Pardo, M. (2009). Dopamine, behavioral economics, and effort. *Frontiers in Behavioral Neuroscience*, 3, 13.
- Salamone, J. D., Correa, M., Mingote, S. M., et Weber, S. M. (2003). Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior : Implications for studies of natural motivation, psychiatry, and drug abuse. *The Journal of Pharmacology and Experimental Therapeutics*, 305(1), 1–8.
- Salamone, J. D., Correa, M., Mingote, S. M., et Weber, S. M. (2005). Beyond the reward hypothesis : Alternative functions of nucleus accumbens dopamine. *Current Opinion in Pharmacology*, 5(1), 34–41.

- Samejima, K., Ueda, K., Doya, et Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, 310(5752), 1337–1340.
- Sanes, J. N. (1986). Kinematics and end-point control of arm movements are modified by unexpected changes in viscous loading. *The Journal of Neuroscience*, 6(11), 3120–3127.
- Sanner, R. M. et Kosha, M. (1999). A mathematical model of the adaptive control of human arm motions. *Biological Cybernetics*, 80(5), 369–382.
- Saunders, J. A. et Knill, D. C. (2004). Visual feedback control of hand movements. *The Journal of Neuroscience*, 24(13), 3223–3234.
- Schmidt, L., Cléry-Melin, M.-L., Lafargue, G., Valabrègue, R., Fossati, P., Dubois, B., et Pessiglione, M. (2009). Get aroused and be stronger : Emotional facilitation of physical effort in the human brain. *The Journal of Neuroscience*, 29(30), 9450–9457.
- Schmidt, L., D’arc, B. F., Lafargue, G., Galanaud, D., Czernecki, V., Grabli, D., Schupbach, M., Hartmann, A., Levy, R., Dubois, B., et Pessiglione, M. (2008). Disconnecting force from money : Effects of basal ganglia damage on incentive motivation. *Brain : a journal of neurology*, 131(5), 1303–1310.
- Schultz, W. (2010). Subjective neuronal coding of reward : Temporal value discounting and risk. *The European Journal of Neuroscience*, 31(12), 2124–2135.
- Schultz, W., Preusschoff, K., Camerer, C. F., Hsu, M., Fiorillo, C. D., Tobler, P. N., et Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 363(1511), 3801–3811.
- Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7), 532–546.
- Scott, S. H. (2008). Inconvenient truths about neural processing in primary motor cortex. *The Journal of Physiology (London)*, 586(5), 1217–1224.
- Searle, L. et Taylor, F. (1948). Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *Journal of Experimental Psychology*, 38(5), 615–631.

- Selen, L. P. J., Beek, P. J., et Dieën, J. H. (2006). Impedance is modulated to meet accuracy demands during goal-directed arm movements. *Experimental Brain Research*, 172(1), 129–138.
- Selen, L. P. J., Beek, P. J., et van Dieën, J. H. (2005). Can co-activation reduce kinematic variability? A simulation study. *Biological Cybernetics*, 93(5), 373–381.
- Selen, L. P. J., Franklin, D. W., et Wolpert, D. M. (2009). Impedance control reduces instability that arises from motor noise. *The Journal of Neuroscience*, 29(40), 12606–12616.
- Sethi, S. P. et Thompson, G. L. (1972). *Applications of Optimal Control Theory in Management Science and Economics*. New York, NY : Springer.
- Shadmehr, R. et Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185(3), 359–381.
- Shadmehr, R. et Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5 Pt 2), 3208–3224.
- Shadmehr, R., Orban de Xivry, J. J., Xu-Wilson, M., et Shih, T.-Y. (2010a). Temporal discounting of reward and the cost of time in motor control. *The Journal of Neuroscience*, 30(31), 10507–10516.
- Shadmehr, R., Smith, M. A., et Krakauer, J. W. (2010b). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89–108.
- Simpkins, A. et Todorov, E. (2009). Practical numerical methods for stochastic optimal control of biological systems in continuous time and space. In *IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (pp. 212–218).
- Soechting, J. F. et Lacquaniti, F. (1983). Modification of trajectory of a pointing movement in response to a change in target location. *Journal of Neurophysiology*, 49(2), 548–564.
- Sozou, P. (1998). On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society B : Biological Sciences*, 265(1409), 2015–2020.

- St Onge, J. R. et Floresco, S. B. (2009). Dopaminergic modulation of risk-based decision making. *Neuropsychopharmacology*, 34(3), 681–697.
- Stengel, R. (1994). *Optimal Control and Estimation*. New York, NY : Dover.
- Stephens, D. et Krebs, J. (1986). *Foraging Theory*. Princeton, NJ : Princeton University Press.
- Stevens, J. R., Rosati, A. G., Ross, K. R., et Hauser, M. D. (2005). Will travel for food : Spatial discounting in two New World monkeys. *Current Biology*, 15(20), 1855–1860.
- Sugrue, L. P., Corrado, G. S., et Newsome, W. T. (2005). Choosing the greater of two goods : Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, 6(5), 363–375.
- Sutton, R. S. et Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA : MIT Press.
- Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., et Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research*, 142(2), 284–291.
- Tanaka, H., Krakauer, J. W., et Qian, N. (2006). An optimization principle for determining movement duration. *Journal of Neurophysiology*, 95(6), 3875–3886.
- Teasdale, N., Philhower, J. G., et Stelmach, G. E. (1990). Temporal movement control in patients with Parkinson’s disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 53(10), 862–868.
- Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9), 907–915.
- Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5), 1084–1108.
- Todorov, E. (2007). Optimal control theory. In K. Doya, S. Ishii, A. Pouget, et R. Rao (Eds.), *Bayesian Brain : Probabilistic Approaches to Neural Coding* (pp. 269–298). Cambridge, MA : MIT Press.

- Todorov, E. et Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11), 1226–1235.
- Trommershäuser, J., Maloney, L. T., et Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, 12(8), 291–297.
- Turner, R. S. et Desmurget, M. (2010). Basal ganglia contributions to motor control : A vigorous tutor. *Current Opinion in Neurobiology*, 20(6), 704–716.
- Turner, R. S., Grafton, S. T., Votaw, J. R., DeLong, M. R., et Hoffman, J. M. (1998). Motor subcircuits mediating the control of movement velocity : A PET study. *Journal of Neurophysiology*, 80(4), 2162–2176.
- Uno, Y., Kawato, M., et Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61(2), 89–101.
- Valero-Cuevas, F. J., Venkadesan, M., et Todorov, E. (2009). Structured variability of muscle activations supports the minimal intervention principle of motor control. *Journal of Neurophysiology*, 102(1), 59–68.
- van Beers, R. J. (2007). The sources of variability in saccadic eye movements. *The Journal of Neuroscience*, 27(33), 8757–8770.
- van Beers, R. J., Baraduc, P., et Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 357(1424), 1137–1145.
- van Beers, R. J., Haggard, P., et Wolpert, D. M. (2004). The role of execution noise in movement variability. *Journal of Neurophysiology*, 91(2), 1050–1063.
- van der Helm, F. et Rozendaal, L. (2000). Musculoskeletal systems with intrinsic and proprioceptive feedback. In J. Winters et P. Crago (Eds.), *Biomechanics and Neural control of Posture and Movement* (pp. 164–174). New York, NY : Springer.
- Van Opstal, A. J. et Van Gisbergen, J. A. (1987). Skewness of saccadic velocity profiles : A unifying parameter for normal and slow saccades. *Vision Research*, 27(5), 731–745.

- Veltkamp, M., Aarts, H., et Custers, R. (2009). Unravelling the motivational yarn : A framework for understanding the instigation of implicitly motivated behaviour resulting from deprivation and positive affect. *European Review of Social Psychology*, 20(1), 345–381.
- Wakker, P. et Kahneman, D. (1997). Back to Bentham ? Explorations of experience utility. *Quarterly Journal of Economics*, 112(2), 375–405.
- Walton, M. E., Kennerley, S. W., Bannerman, D. M., Phillips, P. E. M., et Rushworth, M. F. S. (2006). Weighing up the benefits of work : Behavioral and neural analyses of effort-related decision making. *Neural Networks*, 19(8), 1302–1314.
- Wanat, M. J., Kuhnen, C. M., et Phillips, P. E. M. (2010). Delays conferred by escalating costs modulate dopamine release to rewards but not their predictors. *The Journal of Neuroscience*, 30(36), 12020–12027.
- Watanabe, K., Lauwereyns, J., et Hikosaka, O. (2003). Effects of motivational conflicts on visually elicited saccades in monkeys. *Experimental Brain Research*, 152(3), 361–367.
- Welch, G. et Bishop, G. (1995). *An Introduction to the Kalman Filter*. Technical Report TR95-041, University of North Carolina at Chapel Hill.
- Winter, D. A. (2009). *Biomechanics and Motor Control of Human Movement*. Hoboken, NJ : John Wiley & Sons Inc.
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, 5(6), 483–494.
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–216.
- Wolpert, D. M. et Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3 Suppl, 1212–1217.
- Wolpert, D. M., Ghahramani, Z., et Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882.
- Wong, J., Wilson, E. T., Malfait, N., et Gribble, P. L. (2009). Limb stiffness is modulated with spatial accuracy requirements during movement in the absence of destabilizing forces. *Journal of Neurophysiology*, 101(3), 1542–1549.

- Worringham, C. J. (1991). Variability effects on the internal structure of rapid aiming movement. *Journal of Motor Behavior*, 23(1), 75–85.
- Wyvell, C. L. et Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward : Enhancement of reward "wanting" without enhanced "liking" or response reinforcement. *The Journal of Neuroscience*, 20(21), 8122–8130.
- Xu-Wilson, M., Zee, D. S., et Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, 196(4), 475–481.
- Zajac, F. E. (1989). Muscle and tendon : Properties, models, scaling, and application to biomechanics and motor control. *Critical Reviews in Biomedical Engineering*, 17(4), 359–411.