Behavioral Studies of Dynamical Control Policies Underlying Human Reaching Movements

Antoine De Comité

Thesis submitted in partial fulfillment
of the requirements for the degree of
Ph.D. in Engineering Sciences

February 2022

Thesis Committee:

Prof. Philippe **Lefèvre** (Advisor)	UCLouvain, Belgium
Prof. Frédéric **Crevecoeur** (Advisor)	UCLouvain, Belgium
Prof. Julien **Hendrickx**	UCLouvain, Belgium
Prof. Andrew **Pruszynski**	Western University, Canada
Dr. David **Thura**	Lyon 1 University, France
Prof. Roland **Keunings** (Chair)	UCLouvain, Belgium

Acknowledgements

Il n'est pas loin le temps où, il y a un peu plus de 4 ans, je prenais pour la première fois la route du labo pour entamer cette thèse. J'allais alors découvrir un nouveau domaine de recherche qui m'était jusqu'alors inconnu pour me retrouver 4 ans plus tard totalement épanoui. Ce parcours n'a bien évidemment pas été un voyage solitaire et j'aimerai remercier sincèrement toutes les personnes qui m'y ont accompagnées.

J'ai eu la chance d'avoir deux superviseurs, qui m'ont accompagné et guidé tout au long de ces années. Frédéric, merci pour avoir partagé avec moi ta connaissance de la littérature en pointant sans cesse vers de nouvelles références, pour ton enthousiasme et ton envie d'aller de l'avant qui m'ont aidés à repousser mes limites, et pour les commentaires toujours très pertinents qui m'ont permis d' améliorer ma rédaction. Philippe, merci pour ta patience à toute épreuve quelle que soit la situation, pour ta rigueur scientifique qui m'a permis plus d'une fois d'affiner mon approche et pour ta bienveillance. Je n'aurai pas pu rêver mieux que vous deux pour m'accompagner dans ce travail.

Je tenais aussi à remercier tous les membres de mon jury de thèse pour leur évaluation critique et pertinente de mon travail. Merci au Dr. David Thura, au Prof. Julien Hendrickx, au Prof. Andrew Pruszynski et au Prof. Roland Keunings d'avoir accepté de faire partie de ce jury. J'aimerai également exprimer ma gratitude au Prof. Andrew Pruszynski pour m'avoir accueilli à bras ouverts dans son groupe pour un séjour de recherche. Ce fut pour moi une expérience très enrichissante qui m'a permis de découvrir tant de nouvelles choses et de clotûrer en beauté ces 4 années et demi.

Au cours de toutes ces années, j'ai eu la chance de travailler avec de nombreux collègues irremplaçables. Merci aux membres anciens et nouveaux du SensMotion Lab avec qui j'ai eu la chance de voyager en conférences, de discuter de recherche, de réparer des setups parfois capricieux et surtout de partager tant de bons moments. Je n'oublie pas tous les collègues d'Inma et de Cosy qui sont trop nombreux pour que je puisse les citer. Un grand merci aux membres du Superlab et du Pruszynski lab de m'avoir accueilli à Western, un merci particulier à Jonathan et Olivier pour m'avoir permis d'utiliser le superbe outil qu'ils ont développé. J'aimerai également remercier Sophie et Morgane avec qui j'ai eu la chance de collaborer ainsi que tous les membres de Neuromatch Academy qu'ils soient étudiants, assistants, bénévoles ou organisateurs.

Merci à Marie-Christine, Pascale, Etienne, Julien, Cathy et Leila pour leur aide administrative et technique qui m'a tant aidé durant cette thèse. Merci également aux équipes de Kinarm et Delsys pour leur professionalisme et leur aide précieuse, particulièrement Paul, Duncan et Anne. Sans eux, nombreux résultats de cette thèse seraient probablement encore à l'état embryonnaire.

J'aimerai également remercier mes amis d'ici et d'ailleurs. Tout au long de cette thèse, ils ont toujours été présents et m'ont motivé pour mener à bien cette longue aventure. Un grand merci à Martin, Charles et Nicolas, mes joyeux colocataires, pour les longues soirées où nous refaisions le monde autour d'un verre, à Edouard pour les escapades à pied ou à vélo par tout temps, à tous les membres du Kotangente et associés pour les nombreux souvenirs partagés. Je voudrais remercier sincèrement Vincent pour la relecture, les salades partagées et les pauses café, Louis pour les parties jouées, les discussions interminables et m'avoir appris les secrets des portes logiques, Florimond pour ton enthousiasme envers mes projets divers et variés, les discussions du vendredi soir et nos city trips magic.

Et pour terminer, je souhaiterai remercier du fond du coeur mes parents pour m'avoir donné ce goût pour la science et cette envie d'apprendre tout et n'importe quoi ainsi que mes frère et soeur pour tous les bons moments passés ensemble.

Abstract

Most of our interactions with the world occur through reaching movements. These movements were extensively studied in the past decades and revealed rich mechanisms both during planning and execution. Despite these numerous years of research, very little is known about the entanglement between the planning and execution facets of reaching movements.

This thesis explores the interactions between the planning and execution steps of these reaching movements through a combination of experimental and modeling approaches. First, we demonstrate that the control policies selected to execute movements are not immutable and can be adjusted online to accommodate changes in task demands. These online adjustments suggest the existence of a mechanism whereby changes in the task could elicit replanning of movement. Elaborating on this first finding, we then report that this adjustment can also handle dynamical changes in the task. This finding reinforces the first one as it demonstrates that this adjustment mechanism consists in a continuous feedback loop monitoring task demands to optimally adjust the control policy online. Then, we reveal that selecting more robust control policy, intended at rejecting potential disturbances, reduces flexibility during movement. This last result demonstrates a competition between movement vigor and flexibility.

This work adds to the understanding of the control policies underlying reaching movements by providing new insights into the interactions between movement planning and execution. These results, combined with the modeling work, highlight the fact that movement planning and execution are two facets of a same mechanism underlying reaching movements.

Contents

Chapter I

General introduction

Inspiration is found by looking outward

Oviya Pashiri, sage lifecrafter

The central nervous system is an exquisitely complex and mysterious structure that all animals possess under various forms. From the simple and fully documented nervous system of the *Caenorhabditis elegans* [Cook et al., 2019] to the extremely complex structure of the human nervous system, research has striven, in the past decades, for decoding, understanding and explaining this biological structure. Even though only very few of its functions have been unveiled, it has become an evidence that one of the main purposes of this structure is to allow the host organism to move. Movement is indeed a hallmark of the animal reign, as demonstrated by the annoying fly of the mosquito as one tries to sleep or by the ethereal movements of gymnasts at the Olympics. Although humans are the first observers of the myriad of movements they perform, the reason why these movements are the way they are is still a deeply unanswered question. The present thesis falls within decades of research aiming at explaining how the central nervous system plans and executes movements. More specifically, this work discusses reaching movements performed by humans with their upper limbs by presenting experimental results and comparing the observed behaviors with models borrowed from control theory. As this thesis is at the border between an experimental and a theoretical field, it is primordial to introduce these two research domains. This is

the goal of this first chapter. We first present the current knowledge about reaching movements by reviewing previous experimental results following a chronological order supposed to parallel the temporal unfolding of reaching movements. These experimental results focused on behavioral and physiological data, as the aim of this first section is to draw a portrait of reaching movements by listing their characteristics without discussing their causes. We then discuss the models borrowed from control theory that can be used to simulate and explain reaching movements. This second section sets a theoretical ground to discuss the experimental results presented in this work. The neural implementation of some of these control models within the human brain and the peripheral nervous system will also be discussed in this second section. Finally, the goal of this thesis and the contents of the different chapters will be presented.

1 Experimental studies of reaching movements

In their everyday life, humans perform thousands of reaching movements without even thinking about them. Indeed, any object that they end up holding in their hand is the result of a successful reaching movement whose aim was to reach and grasp that specific object. In the first section of this chapter, we dissect these movements by describing the different steps composing them and detailing how each of them is influenced by various parameters. For this purpose, we refer to previous behavioral studies in humans, non-human primates, and other animal models spanning from the pioneering work of Bernstein [Bernstein, 1967] to the most recent studies.

This section follows the chronological order of the successive processes occurring when the agent[1] executes a reaching movement. The first process in this temporal unfolding is *movement planning*, which can be itself split in different parts. During movement planning, the agent decides which behavior to adopt for the task they have to execute based on the information they collected through their sensory inputs.

[1] In this thesis, the term *agent* refers to the entity that takes decision and execute action. It can refer to the human or animal performing the experimental task or to the mathematical model simulating movements.

We first describe in section 1.1.1 how the agent can use their sensory inputs to collect information about their environment and their own body. Then we describe how this sensory information can be optimally combined with cognitive information, described in section 1.1.2, within the decision process, covered in section 1.1.3, selecting the agent's behavior for the next step. This second step is *movement execution* which consists in the dynamical application of the behavior selected by the agent to execute the task optimally according to some criterion. In this second part, we start by discussing the specificity of the behavior selected by the agent in unperturbed conditions (section 1.2.1), and then we expand it to an environment involving visual (section 1.2.2) or mechanical perturbations (section 1.2.3) of the system. It is worth noting that movement planning and execution are not two sequential processes and that they can happen in parallel (see section 1.3 for more details). Figure I.1 schematically summarizes the structure of this section with respect to the temporal unfolding of the processes underlying reaching movements.

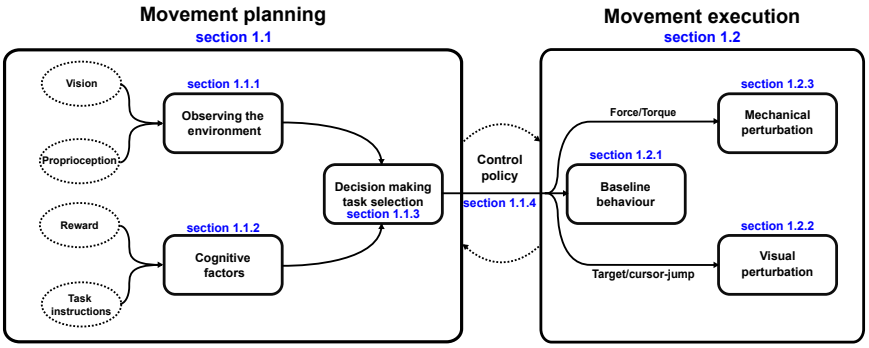

Figure I.1: Temporal unfolding of the different processes underlying reaching movements. The first step, *movement planning* consists in selecting a control policy by integrating, within the decision process, the observation of the environment performed through sensory inputs and cognitive factors such as reward and task instructions. The control policy is then used during *movement execution* in both unperturbed and perturbed conditions (either with mechanical or with visual perturbations). The sections of the text associated with the different part of this structure are also indicated in this figure.

1.1 Movement planning

The first step that the agent has to go through when they decided to execute a reaching movement is movement planning[2]. Movement planning is primordial as it aims at selecting the agent's behavior during movement for any possible limb position while integrating task-related information collected through sensory modalities and other external inputs. The first source of task-related information that is discussed herebelow is the one mediated by vision an proprioception (section 1.1.1). The second source that is discussed concerns the higher-level information conveyed to the agent either by the mean of specific task rules or implicit and explicit reward (section 1.1.2). Task-related knowledge collected through these two modalities is then integrated during a decision process aiming at selecting the agent's behavior (section 1.1.3).

1.1.1 Observation of the environment

A complete and detailed understanding of the initial state of the environment and of the agent's body is primordial to select the behavior to adopt during movement. This understanding first requires to collect information through the sensory inputs relevant to the context: *proprioception* and *vision*.

Proprioception

Proprioception is the channel whereby information about the limbs position and movement is conveyed to the brain and the peripheral nervous system. Since the pioneering works by Mott and Sherrington [Mott and Sherrington, 1894], proprioception is known to be a key element in the estimation of the postural state of the motor apparatus as it conveys information to which other senses are blind. It is mainly mediated by three types of receptors, respectively located under the skin, within the muscles belly and in the tendons: *touch receptors*, *muscle spindles*, and *Golgi organs* respectively. *Skin deformations* convey information to the brain

[2]Here, the nomenclature introduced by Trommershaüser [Trommershäuser et al., 2003] will be adopted and this first step will be referred to *movement planning*. In contrast to *motor planning* which is concerned in defining and finding a deterministic sequence of motor commands, movement planning has the flexibility to determine a global behavior whatever the state of the limb is with respect to the goal.

about the state of its limb. The skin is indeed populated with many different kinds of sensors (see [Johnson and Hsiao, 1992, Johnson et al., 2000, Johnson, 2001] for reviews) sensitive to various types of stimulus such as identations or vibrations for instance [Johansson and Westling, 1988, Johansson and Westling, 1984]. A marvellous characteristics of these captors is that they can work in concert using each and everyone's sensitivity to encode these various stimuli [Delhaye et al., 2018, Saal and Bensmaia, 2014, Weber et al., 2013]. Skin deformation and touch are not only used during grasping and object manipulation, they are also important for reaching movements. Indeed, previous work demonstrated that the sensorimotor control of upper limb reaching movements and the sense of touch are closely related [Crevecoeur et al., 2016, Retschechtko and Pruszynski, 2020, Forgaard et al., 2021, Avraham and Nisky, 2020, Hernandez-Castillo et al., 2020]. *Muscles spindles* are composed of two types of afferents grouped in bundles within the muscle fibers (see [Hulliger, 1984] for review). These spindles are sensitive to the size of muscle length change and to its speed [Matthews, 1972, Proske and Gandevia, 2009, Cheney and Preston, 1976, Edin and Vallbo, 1990a] during passive movements, which implicitly also conveys information about the limb movement and speed. Their activities during active movements are more complex than the passive ones and still poorly understood [Dimitriou and Edin, 2008a, Dimitriou and Edin, 2008b]. Moreover, these muscle spindles are also known to be sensitive to muscle vibrations [Burke et al., 1976, Roll et al., 1989] which could elicit illusion of limb movement when these muscle spindles are artificially vibrated [Goodwin et al., 1972]. *Golgi tendons organs* are located in the tendons and are sensitive to muscle tensions and therefore can convey information about the forces and torques applied by and to the different joints [Mileusnic and Loeb, 2009, Edin and Vallbo, 1990a, Edin and Vallbo, 1990b, Matthews, 1933]. The combination of all these sensitive organs provide an exquisitely rich knowledge about the state of the agent's limb. This is however not enough to properly plan movement as proprioception alone does not provide any information about the task environment.

Vision

Vision is a complementary (and sometimes redundant) source of information that with proprioception allows for a full comprehension of the state of the body and environment. Vision is indeed complementary to proprioception as it provides information about parameters to which proprioception is blind (e.g. targets or obstacles in the environments, visual cues related to the task, ...). Since the demonstration by Woodworth that movements are more accurate with vision than without [Woodworth, 1889], it is evident that vision plays a major role in movement planning. Indeed, it was demonstrated that vision of the hand at the beginning of movement causes more accurate movements than when not available [Desmurget et al., 1997, Desmurget et al., 1998, Prablanc et al., 1979b]. Moreover, when the vision of the initial position of the limb is distorted, the movements exhibit systematic errors [Rossetti et al., 1995, Sainburg et al., 2003]. In addition to this information about the initial state of the limb, vision also conveys information about the different targets and obstacles located in the environment. The term *vision* refers to all the information collected by the visual system whether it is retinal processing of visual stimuli or information related to the eye and head orientation. The retinal information provides accurate estimation of the target position if this one is located in the fovea[3]. Indeed, the encoding of the visual stimulus is degraded if it falls in the peripheral visual field [Paillard and Amblard, 1985, Westheimer, 1984]. This degraded encoding results in an increase of the end-point error during hand pointing tasks [Bock, 1993, Bock, 1986, Roll et al., 1986, Prablanc et al., 1979a]. Moreover, it has been demonstrated that the estimation of the location of the visual stimulus is improved when the retinal information is combined with concomitant extra-retinal sensory inputs [Prablanc et al., 1979b, Blouin et al., 1995, Bock, 1986] such as the orientation of the eye in the eyeball. The orientation of the head also have a similar impact: the estimation of the location of the visual stimulus is improved with concomitant head orientation [Brotchie et al., 1995]. Interestingly, the eye and head orientation also exhibit more accurate estimation for non peripheral stimuli [Biguer et al., 1984, Vanden Abeele et al., 1993, Rossetti et al., 1994, Paillard and Amblard, 1985]. By combining all the information col-

[3]The fovea is a part of the retina that is densely populated with rods and cones and is therefore associated with a higher visual acuity than the rest of the retina

lected through vision, the agent is able to collect information about the limb and the different targets, complementary to what proprioception collected.

Multi-sensory integration

Vision and proprioception are not processed individually but instead are integrated as a whole through a combined neural and cognitive mechanism [Graziano and Gross, 1993, Driver and Spence, 1998, Spence et al., 2000, Doyle and Walker, 2002]. This combination of different sensory modalities is referred to as *multi-sensory integration*. In this work, we are interested in upper limb movements and more specifically reaching movements. An extremely large body of work investigated multi-sensory integration in upper limb movements and demonstrated that the central nervous systems exquisitely combines various sensory inputs such as vision, proprioception, sound, and even olfaction [Klatzky et al., 1987, Klatzy et al., 2000, Klatzky and Lederman, , Jenmalm et al., 2000, Zhariev and MacKenzie, 2007, van Beers et al., 2002, Patchay et al., 2003, Patchay et al., 2006, Sober and Sabes, 2003, Aziz-Zadeh et al., 2004, Gazzola et al., 2006, Castiello et al., 2006, Camponogara and Volcic, 2019] .

It is hypothesized that the brain integrates these different sensory modalities in a statistically optimal way [Angelaki et al., 2009, Kording et al., 2007]. This optimality means that the estimated variable is the most reliable value for this encoding in terms of maximum likelihood estimation. According to Bayes' rule, the optimal combination of these different sensory inputs is a linear combination of them, weighted by their relative precision[4]. This signifies that the estimate will be closer to the sensory input associated with the smallest variance. This optimal sensory cue combination is not only good at weighting different sensory inputs based on their precision but it can also factor different delays for the different sensory inputs [Cluff et al., 2015]. Many experimental evidence support this optimal cue combination in different laboratory tasks. Ernst and Banks [Ernst and Banks, 2002] demonstrated in their seminal studies that weight and haptic feedback were combined optimally to estimate the size of an object, demonstrating that the brain was able to perform opti-

[4]In statistics, the precision of a random variable is defined as the inverse of its variance

mal cue integration. Kording and Wolpert [Körding and Wolpert, 2004] demonstrated that humans participants were able to learn prior distributions and tune their behavior accordingly in reaching experiments followed, by other reaching tasks reporting congruent results [Trommershäuser et al., 2005, Wolpert and Landy, 2012]. The multisensory integration of proprioception and vision is a central concept in reaching movements because it does not only matter for determining the initial state of the limb but also estimate the state of the limb at any time during movement. However, some questions regarding this multi-sensory integration of signals with asynchronous delays are still unanswered [Cluff et al., 2015]. The sensory delays associated with proprioception and vision indeed play an important role in the ability of the agent to properly respond to visual and mechanical perturbations [Cameron et al., 2014].

1.1.2 Cognitive factors

In contrast to the parameters developed in section 1.1.1, some task-related factors cannot be directly interpreted by the senses and require other kind of processing before being integrated within the selection of the agent's behavior. These factors will be referred to as *cognitive factors* and typically correspond to information that require cognitive processing that the sensorimotor system alone is not able to perform. In this section, we first discuss specific task rules and then the explicit or implicit task reward.

Task rules

Specific instructions provided to the agent have a primordial influence on the selection of the control policy. Indeed many behavioral experiments have revealed that the agent's behavior and specifically the way they handle unexpected perturbations (see section 1.2.3) depends on the instructions they were given. Classical experimental paradigms that modulate the task instructions are the GO/NO-GO and RESIST/LET GO paradigms, which have been demonstrated to modulate long latency feedback responses [Calencie and Bawa, 1985, Capaday et al., 1994, Colebatch et al., 1979, Crago et al., 1976, Rothwell et al., 1980, Shemmell et al., 2009]. These specific task instructions can also be more subtle and can dictate the agent's interactions with the environment and its different tar-

gets or obstacles. It has indeed been demonstrated that humans are able to integrate specific instructions relative to obstacles in the selection of their control policy [Sabes and Jordan, 1997, Nashed et al., 2012, Cross et al., 2019]. This body of work confirms that specific task rules and instructions modulate the selection of the appropriate control policy.

Reward

Another cognitive factor that influences movement planning is the reward associated with the task and with each of the available targets. Target reward is a parameter that has been investigated in many experiments as this is a factor that is relevant to motor adaptation and motor skill learning [Abe et al., 2011, Galea et al., 2015, Wächter et al., 2009]. In this thesis, we will focus on its role on the motor policies used by humans to execute movement, and let its role on motor adaptation and motor skill learning aside.

The impact of reward on the control of saccadic eye movements has been investigated in many experimental studies on humans and monkeys. Positive reward invigorates movement as demonstrated by the increase in movement speed and the decrease in reaction time when reaching to target associated with the larger reward [Manohar et al., 2015, Manohar et al., 2017]. Interestingly, similar behaviors have been reported in experimental paradigms where the target reward was implicitly modulated, for instance by providing more informative target such as images of faces [Xu-Wilson et al., 2009].

A similar impact of target reward had been observed in reaching movements where an increase of movement vigor has been also reported for higher target reward in both animal [Mosberger et al., 2016, Opris et al., 2011] and human studies [Summerside et al., 2018, Esteves et al., 2016]. Since reward modulates movement speed, it also modulates movement duration. This modulation of movement duration has been demonstrated to follow an hyperbolic law, suggesting an hyperbolic temporal discounting of reward [Shadmehr et al., 2010a, Haith et al., 2012]. All these results demonstrated that target reward, whether it is explicit or implicit, modifies the control policies used by humans to execute movement and is therefore a parameter that cannot be neglected when investigating movement planning.

Motor cost and effort

During movements of large amplitude, the agent has to take factors such as the arm anisotropy, inertial passive forces or muscle viscoelasticity into account as they became non negligible [Gordon and Ghez, 1987, Flash, 1987, Vindras et al., 2005, Guigon et al., 2007]. This means that movement planning also integrates intrinsic limb properties [Soechting and Lacquaniti, 1981, Kaminski and Gentile, 1989] and even some knowledge about the limb dynamics [Uno et al., 1989]. This was proven experimentally in a series of studies conducted by Sabes and colleagues [Sabes et al., 1998, Sabes and Jordan, 1997]. They demonstrated that trajectories adopted to move around obstacles are such that they minimize the impact of potential perturbations by exploiting knowledge about dynamical limb factors.

Besides the knowledge of intrinsic limb properties, the agent also has to take the motor cost into account. Indeed, moving one's arm is not energetically free and there is a cost incurred with it as shown by the impact of motor cost on some motor control aspects in humans [Shadmehr et al., 2019, Huang et al., 2012, Cos et al., 2011] or animals [Alexander, 1996]. Motor cost or effort has a clear impact on the control policies selected to perform movements and it has been used as a factor to determine movement duration and vigor in some motor control theories [Shadmehr et al., 2016, Wong et al., 2021, Rigoux and Guigon, 2012]. Similarly, the framework of Optimal Feedback Control (developed in section 2.2) considers motor cost as one of the main parameters taken into account to determine the control policy. This motor cost can be evaluated in many different ways : muscular force [Kolossiatis et al., 2016], mechanical work [Alexander, 1997], square of the activated force [Nelson, 1983, Ma et al., 1994], or even the squared integral of the torque derivatives [Uno et al., 1989].

1.1.3 Decision-making

In sections 1.1.1 and 1.1.2, we reported that both low-level sensorimotor (the information conveyed by vision and proprioception) and high-level cognitive factors are integrated in the selection of the agent's behavior when only one target is available. However, in everyday life situations, the agent often faces situations were more than one target is available. The goal of this section is to discuss the decision process that the agent uses in such situations, and to infer how the low- and high-level factors

are integrated in this decision-process. In this section, we first describe the *drift-diffusion models* as they were a starting point to many of the experimental studies discussed later. The second part reviews decision-making processes that are somehow related to reaching movements. Finally, the interesting interplay between low- and high-levels factors in these decision process will be discussed.

Drift-diffusion models

The *drift-decision models* are a family of models introduced to explain and infer the decision process that takes place when an agent is faced with two or more alternatives[5] [Laming, 1968, Ratcliff, 1978]. These models exploit the sequential probability ratio test (SPRT) developed by Wald [Wald, 1945, Wald and Wolfowitz, 1948] to model decision process that occurs through a competition between the two alternatives whose respective evidences are modeled by a gradual accumulation following a random walk[6]. More specifically, in the binary decision process, the agent has to select between the alternatives A and B, each of them being associated with an accumulation variable x_A and x_B. Throughout the temporal unfolding of the process, the agent is collecting bits of information about each hypothesis, δx_A and δx_B, respectively, that sums up with the corresponding accumulation variables. When one of the two alternatives reaches a predefined threshold, the associated decision is taken. A schematic representation of these drift-diffusion models is presented in Figure I.2, where the agent had to decide whether the patch of points is moving to the left or to the right (Figure I.2A). Assuming that the agent is looking at the patch for a long time, the variables representing the evidences are represented in figure I.2B.

Interestingly, the principle of evidence accumulation underlying drift-diffusion models prompted the experimental search for the existence of such an accumulation neural signal in the brain. Indeed, if the brain is using a decision process similar to the drift-diffusion models, it should somehow accumulate evidence, and some gradually increasing signals might be observed in the brain. These accumulation signals were indeed

[5]The models were initially developed for binary decision processes but were eventually expanded to processes where a finite set of alternative outcomes are possible.

[6]The term *random walk* here signifies that, at every time step, the accumulated evidence is defined by $dx = Adt + \omega$ where A is the evidence of the current time step and $\omega \sim \mathcal{N}(0, 1)$ follows a normal distribution.

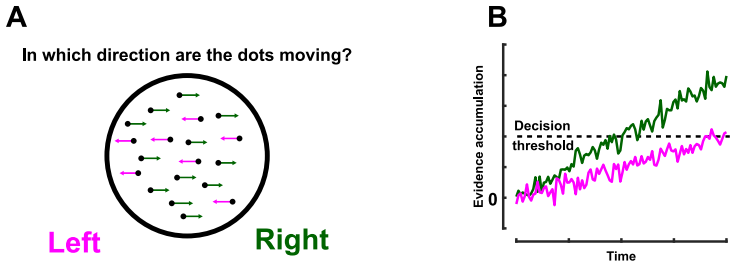

Figure I.2: Representation of the drift-diffusion models. **A**. The task consists in determining the direction towards which a patch of dots is moving based on the individual dots movements. Participants are instructed to make a decision when they are confident enough in their choice. **B**. Schematic representation of the evidence accumulation for both alternatives as predicted by the drift-diffusion models. Both variables started at zero and the first one to cross the decision threshold is considered the winner. In this case, the agent answers that the patch of dots is moving to the right.

found in various experimental paradigms (see [Gold and Shadlen, 2007] for review) suggesting that the animal and human central nervous systems could use mechanisms similar to these drift-diffusion models.

Decision-making models for reaching movements

Decision-making is a central mechanism in sensorimotor control which does not only take place when there are multiple targets available to the agent. Indeed, even when only one target is presented, there is a multitude of behaviors that the agent can select to execute this movement (different speeds, trajectories, ...). As stated by Wolpert and Landy [Wolpert and Landy, 2012], movement planning is thus all about decision-making. In this section, we will focus on the decision process that takes place when several alternatives are available, and selection between reaching policies to a same target will be briefly discussed in section 1.1.4. The reason why decision-making was covered in this thesis is that the experimental work developed here investigates the potential entanglement of movement planning and execution. Therefore, since movement planning is mainly concerned with decision making, it is of the highest importance to cover that topic in this thesis introduction.

When the agent has to select an option between two potential targets, their decision is influenced by both low- and high-level factors discussed above. Indeed, it has been demonstrated that partic-

ipants tend to favor targets associated with the smaller biomechanical costs [Cos et al., 2011, Burk et al., 2014, Cos et al., 2014]. This suggests that, somehow, a representation of the cost associated with the movements to the different targets is computed and that these values are compared to each other. Similarly, the agent's decision reflects cognitive factors, as demonstrated by the influence of reward and punishment on the selection between the different options [Trommershäuser et al., 2003, Trommershäuser et al., 2008]. Trommershauser and colleagues [Trommershäuser et al., 2008] demonstrated that participants were able to integrate abstract cognitive factors such as reward or penalty and select the movement that will maximize their outcome, at very short response times.

Up to now, the low- and high-level sensorimotor factors where considered separately for the decision-making process. However, in real-life both factors can be involved in the same decision process. This gives rise to the interesting question of their integration in the decision making process. This question was addressed by Cisek [Cisek, 2012], that suggested an interplay between these two types of parameters. In this review, Cisek suggested that the decision between actions (e.g. going to the left or right target) is not only determined by the sensorimotor representations of the different options but also integrates abstract representations of the actions outcome (e.g. how much reward is associated with each target). This theory was further explored by Rigoux and Guigon [Rigoux and Guigon, 2012] who presented a model that, similarly, supports the interplay between cost (associated with low-level sensorimotor factors) and reward (associated with high-level cognitive factors) in the decision-making processes.

Besides these sensorimotor and cognitive factors, the decision process can also be influenced by the urgency which the participants must make decisions with. In this case, in the framework of *drift-diffusion models*, the decision will be taken before the threshold is reached. The theoretical framework associated with this phenomenon is the so-called *urgency-gating model* [Cisek et al., 2009]. This model proposes that the buildup of neural activity reported in previous experiments could be linked to a growing signal representing the urgency to respond, instead of the temporal integration of evidence [Cisek et al., 2009, Thura et al., 2012]. This urgency-gating model was further investigated in a token task where humans and non-humans

primates had to make decisions, and neural correlates from the motor and premotor cortices confirmed the urgency-gating hypothesis [Thura and Cisek, 2014]. These results supported the hypothesis that the urgency was also a critical factor that has to be taken into account in the investigation of the decision processes.

The decision processes that we consider in the present thesis and those modeled by the *drift-diffusion models* are processes that take time. Indeed, the decisions are not instantaneously taken at the beginning of the decision process. This temporal dimension of decision making gives rise to an interesting tradeoff between the accuracy of the decision and the time required to reach that decision. This tradeoff, called *speed-accuracy tradeoff* (SAT) has been demonstrated to happen in many different experimental paradigms involving vision, olfaction, audition or even memory [Reed, 1973, Wickelgren, 1977, Reddi and Carpenter, 2000, Reddi et al., 2003, Palmer et al., 2005, Rinberg et al., 2006, Ings and Chittka, 2008, Bogacz et al., 2010]. The underlying principle of this speed-accuracy tradeoff is that a decision cannot at the same time occur quickly and be accurate. Besides the behavioral support of this SAT, many studies demonstrated neural evidence of this tradeoff in various brain area. For instance, it has been shown that such signals exist in the monkey brain when they execute a perceptual discrimination task in the frontal eye field [Hanes and Schall, 1996, Kim and Shadlen, 1999, Woodman et al., 2008, Ding and Gold, 2012], in the lateral intraparietal area [Roitman and Shadlen, , Gold and Shadlen, 2007, Hanks et al., 2014] or in the superior colliculus [Horwitz and Newsome, 1999, Ratcliff et al., 2007]. Interestingly, the existence of this tradeoff indicates that the decision processes can be interrupted at any time and that an associated decision will be available. This point is critical to the decision we conduct in section 1.3.

1.1.4 Control policy

Movement planning and the different steps detailed above result in the selection of a control policy. This control policy is selected by the agent and defines their behavior during movement. Conceptually, a control policy is a mapping between the state of the system (i.e. the agent's body and environment) and the actions that the agent takes (i.e. the motor commands), which can be mathematically expressed as follows:

$$\pi\left(\mathbf{x}\left(t\right),t\right)=\mathbf{u}\left(t\right), \tag{I.1}$$

where $\mathbf{x}\left(t\right)$ is the state vector, which is a set of variables that unequivocally describes the system, and $\mathbf{u}\left(t\right)$ is the action selected by the agent given that state of the system.

This definition of control policy is central to this thesis and is worth to be discussed in details. We reviewed that the agent's behavior, governed by the control policy, depends on both low- (see section 1.1.1) and high-level factors (see section 1.1.2). This means that, for a given system state $\tilde{\mathbf{x}}\left(t\right)$ (i.e. joint configuration and velocity), the motor command selected by the agent in two different tasks, $\mathbf{u}_1\left(\tilde{\mathbf{x}}\left(t\right),t\right)$ and $\mathbf{u}_2\left(\tilde{\mathbf{x}}\left(t\right),t\right)$, differs if and only if one task-defining parameter differs. This selected control policy impacts the *baseline* control (see section 1.2.1) characterized by voluntary unperturbed movements to the target goal but also the *feedback* responses to perturbations, as will be discussed in section 1.2. In sections 1.1.1 and 1.1.2, we described different agent's behaviors induced by modulation of different task-related parameters. These differences are actually attributable to different control policies associated with the different conditions. The control policy, selected during movement planning, is responsible for the behaviors observed experimentally as it dictates the agent's action for any state of the limb.

1.2 Movement execution

Once the control policy has been selected by the agent during movement planning, it can be applied to execute movement. In this section, we review previous studies that investigated the influence of the different factors developed above on animals and humans behavior during reaching movements. In section 1.2.1, we discuss the baseline behavior without perturbations. Then, in sections 1.2.2 and 1.2.3, we review the paradigms where visual and mechanical perturbations were respectively used and how they conveyed information about the control policy. Perturbation paradigms are primordial to investigate the impact of various factors on these control policies. Indeed, perturbing the system gives the opportunity to put it in an unexpected state which elicits a feedback[7] response to the perturbation.

[7]By definition, a feedback response is a response (ie. an action of the agent) that integrates the unexpected state that the system has reached after a perturbation.

1.2.1 Baseline behavior

The application of the control policy in absence of any kind of perturbation yields to the so-called *baseline behavior*. Despite the absence of perturbations, this baseline behavior already provides a huge amount of information regarding the control policy used to perform movements.

A hallmark of this baseline behavior is its ability to exploit the large number of degrees of freedom inherent to the limb dynamics to perform movements using invariant kinematic and dynamic features, which gives rise to stereotypical behaviors. This property has been demonstrated for both 2- and 3-dimensional reaching movements in various paradigms [Flash and Hogan, 1985, Morasso, 1981, Soechting and Lacquaniti, 1981, Biess et al., 2007, Atkeson and Hollerbach, 1985, Hollerbach and Flash, 1982, Bernstein, 1967, Wolpert et al., 1995, Haggard et al., 1995]. For instance, there is an inverse non-linear relationship between the tangential hand speed and the curvature of the trajectory during curved movement as stated by the *two-thirds power law* [Lacquaniti et al., 1983], illustrated in Figure I.3. These relationships between hand position and hand velocity across movements and paradigms reveal that the central nervous system somehow exploits the tremendous number of degrees of freedom inherent to the upper limb in a stereotyped way to execute movement.

The control policy used by humans to execute movement is also sensitive to target redundancy and exploits this redundancy when available. This is demonstrated by the modulation of the end-point distribution of hand position with the target width : wider distributions were associated with wider targets [Berret et al., 2011, Knill et al., 2011, Nashed et al., 2012]. These results are congruent with the claim that the control policy used for reaching exploit limb redundancy because they demonstrated that the central nervous system has an efficient way of dealing with redundancy associated with the task. This exploitation of the task redundancy might be explained by the way the control policy is selected (see 2.2 for more details).

Cognitive factors such as reward also have a direct effect on the behavior in this baseline condition as it has been shown that humans tend to move faster towards more rewarding targets, as highlighted by faster saccadic eye movements towards targets associated with higher util-

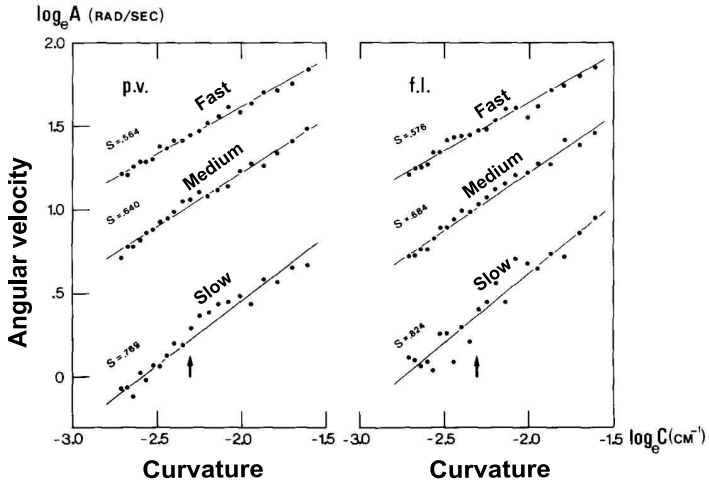

Figure I.3: Representation of the *two-thirds power law* obtained for two different subjects drawing at different speeds (fast, medium and slow). The angular velocity of their movements was correlated with their curvatures in a stereotyped way, whatever the velocity was. Adapted from [Lacquaniti et al., 1983].

ity or monetary reward [Xu-Wilson et al., 2009, Manohar et al., 2015]. Similar results have been reported in upper limb reaching movements [Esteves et al., 2016, Summerside et al., 2018], where more vigorous movements were observed towards more rewarding targets.

However, baseline behavior alone is not enough to grasp the interesting complexity of the reaching control policies. In the next two sections, we discuss two types of non-invasive perturbation paradigms that provide new insights in these control policies.

1.2.2 Visual perturbations

So far, the reported experimental findings about movement execution indicated that the control policy is modulated by low-level sensorimotor and high-level cognitive factors in a feedforward way. This means that different control policies are selected for different task prior to movement onset. However, a potential effect of these parameters on the feedback mechanisms within the same movements has not been discussed yet. The goal of this section is to unveil the role of vision on this feedback mechanism and to demonstrate that the feedback part of the reaching control policy is also affected by these factors.

In section 1.1.1, we already discussed the primordial role of vision to understand the task and select the appropriate control policy. Vision is also a key sensory input during movement as it conveys information about potential visual perturbations occurring in the environment. These visual perturbations can be used to elicit corrections that should provide information about a putative modulation of the feedback mechanisms by the parameters discussed above.

A first experimental paradigm that can be used to investigate the agent's response to visual changes is TARGET JUMP. In this paradigm, the target toward which the agent initiated their movement unexpectedly jumps to another location. The first thing that this paradigm, named DOUBLE-STEP PARADIGM in Sarlegna & Mutha's review [Sarlegna and Mutha, 2014], reveals is that these target jumps elicit online motor corrective responses [Megaw, 1974]. These responses do not occur immediately after the target jump: their latency has been reported to be about 100ms [Georgopoulos et al., 1981, Soechting and Lacquaniti, 1983]. The occurrence of these responses to visual perturbation confirmed the existence of a feedback mechanism that relies on vision. An interesting question is whether some parameters that modified movement planning also modified these feedback responses to visual perturbation. The state of the system at perturbation onset influences the agent's response as shown by differences in responses when the perturbation occurred at different time within the movement [Barret and Glencross, 1989, Brenner and Smeets, 1997, Ma-Wyatt and McKee, 2007] or by the modulation of these feedback responses depending on the state of the muscles at perturbation onset [Carlton and Carlton, 1987]. The modulation of these feedback responses with the state of the system at perturbation onset therefore confirms the hypothesis of a state-feedback[8] mechanism used during motor execution and that uses vision as a relevant sensory input.

Higher level cognitive factors also have an impact on these feedback responses to perturbations. Indeed, these responses were shown to be modulated by the task instructions [Cameron et al., 2009, Striemer et al., 2010] or by the expected reward associated with the target [Boulinguez and Nougier, 1999].

[8]State-feedback mechanisms refers to a feedback loop that provides the controller with the state of the system or its estimation.

Altogether, these findings demonstrated that vision of the limb and the environment provide the agent with information about the state of the limb (position and velocities) and about the task (relative position, instructions). The agent collects this information through vision to provide an appropriate feedback response to the perturbation in about 100ms. An interesting point is that these online feedback corrections were also observed when participants were not aware of their occurrence [Goodale et al., 1986, Pélisson et al., 1986] which suggests that this mechanism could be automatic.

1.2.3 Mechanical perturbations

Vision is not the only sensory input that the agent can use to collect information about the state of their limb. As stated in section 1.1.1, proprioception only provides information relative the agent's limb (position, velocity, torques, ...) and is therefore mainly used when a perturbation occurred on that limb. Since it conveys information about the force and torques at various joints, proprioception is particularly useful in paradigms involving mechanical perturbations applied to the agent's hand while they were moving. Indeed, the agent has access to a strictly better estimate of the force or torque applied by the environment by using proprioception rather than vision.

Feedback responses to mechanical perturbations mainly differed from those triggered by visual perturbations in their latencies. Indeed, delays of about 100ms have been reported for visual perturbations [Georgopoulos et al., 1981, Soechting and Lacquaniti, 1983] with the fastest EMG responses that can occur in about 90ms [Pruszynski et al., 2010]. Mechanical perturbations can elicit faster EMG responses in as little as 20-25ms following perturbation onset [Weiler et al., 2019, Pierrot-Deseillignly and Burke, 2005]. Three different types of feedback responses to mechanical perturbations are defined based on the timing from perturbation onset: the short-latency response (25-45ms), the long-latency response (50-100ms), and the voluntary response (100-180ms). Each of these responses has its own sensitivity and can be modulated by different factors as discussed in the subsequent paragraphs.

The short-latency responses transit by the spinal cord and can be modulated in a very limited set of experimental conditions such as specific joints configuration [Weiler et al., 2019], transitions between posture and movement [Thompson et al., 2009], cyclic actions [Dufresne et al., 1980, Duysens et al., 1993, Capaday and Stein, 1986], or volitional intent after extensive training [Mortimer et al., 1981, Wolpaw, 1983]. In most other cases, short-latency responses cannot be modulated experimentally as the investigated parameters are probably handled by cortical circuits that are not involved in these short-latency responses. For instance, an absence of changes in these short-latency responses has even been reported in studies investigating task instructions [Rothwell et al., 1980, Crago et al., 1976, Capaday et al., 1994] or modulation of the target structure [Knill et al., 2011, Nashed et al., 2012, Lowrey et al., 2017] even though these studies reported differences in the long-latency responses.

In contrast to the short-latency responses, the long-latency feedback responses to mechanical perturbations involve a transcortical neural pathway and exhibit modulation by many parameters that also influence baseline behavior. For instance, these feedback responses have been shown to depend on the verbal instructions [Capaday et al., 1994, Colebatch et al., 1979, Crago et al., 1976, Rothwell et al., 1980], the knowledge of the physics of the limb [Kurtzer et al., 2008, Lacquaniti and Soechting, 1984, Lacquaniti et al., 1986, Maeda et al., 2018, Maeda et al., 2020], the intersegmental coupling between elbow, wirst and shoulder [Weiler et al., 2015, Weiler et al., 2016, Maeda et al., 2017], the urgency to respond to the perturbation [Crevecoeur et al., 2013, Dimitriou et al., 2013], the position and structure of the goal target [Nashed et al., 2012, Lowrey et al., 2017, Pruszynski et al., 2008, Knill et al., 2011]. These long-latency responses therefore consist of complex feedback responses to the mechanical perturbations that could integrate both low-level sensorimotor factors and high-level cognitive factors. They differ from the short-latency by their complexity and their amplitude, as they exhibit larger values. They also differ from the voluntary responses in that they occur at shorter latencies and therefore only recruit automatic mechanism that do not interact with the voluntary intent which starts being factored in 100ms after perturbation onset.

Short- and long-latency feedback responses are referred to as *reflex responses* as they consist in automatic responses and do not involve any voluntary intent which is not the case for voluntary responses. These voluntary responses happen a long time after perturbation onset as they necessarily involve an additional process and could integrate roughly any parameters that the short- and long-latency could but in a way that might be counter-intuitive and unpredictable because of voluntary intent.

1.3 Planning and execution, two independent processes

The description of reaching movement developed in sections 1.1 and 1.2 might sound as if movement planning and execution were sequential. Indeed, we first presented movement planning as the process responsible for the selection of the control policy and then explained that movement execution is the process whereby that control policy is continuously applied. In the present section, we want to insist on the fact that these two processes are not two sequential processes as the latter can happen even if the former is not completed yet.

In section 1.1.3, we detailed the *drift-diffusion model* characterized by an accumulation of evidence towards an alternative in a decision making task. This accumulation of evidence is a continuous process which means that the decision made by the agent is built gradually during that time as demonstrated by the speed-accuracy tradeoff underlying any decision (see [Heitz, 2014] for review).

This decision process that can be probed at different time points is not exclusive to the perceptual decision making described in section1.1.3. Indeed, it can be expanded to motor control tasks. As we described in the section 1.1.3, selecting a control policy is also a decision making process. This means that it can be probed while it is unfolding to investigate the potential relationship between cognition (eg. decision-making) and action (eg. motor control). A first series of studies investigated whether an action plan already exists during that decision process in oculomotor task in non-human primates [Gold and Shadlen, 2000, Gold and Shadlen, 2003, Horwitz et al., 2004, Kustov and Robinson, 1996, Connolly et al., 2009]. They reported that the superior colliculus of these non-human primates is involved in both

cognition and action. Moreover, they demonstrated that the action plan used for the oculomotor task was already being prepared during the decision process.

Similar results were found in upper limb motor tasks. Selen and colleagues performed a task where participants had to decide which target to reach between two possibilities based on a random dot motion task [Selen et al., 2012]. They investigated the ongoing decision process by perturbing participant's hand before the end of the decision and revealed that the feedback gains where modulated with time. This result demonstrates that a continuous process mirroring the decision process rules the computation underlying sensorimotor control. Another study leveraged the time delay between the target presentation and the mechanical perturbation to demonstrate comparable build-up in control policy [Yang et al., 2011].

To sum up, movement planning and execution differed in the role they play in sensorimotor control but they can overlap when circumstances force a subject to take an action before all the parameters describing the task context are known.

2 Modeling reaching movements

Modeling is a powerful tool to understand the complexity of a phenomenon and to potentially predict the outcome of future experiments. Many experimental research fields have their modeling counterpart and those are complementary as sometimes the model predicts the future experimental observations (e.g. the existence of the Higgs' boson was predicted before being observed) and sometimes an unexpected experimental result triggers a whole field of modeling research (e.g. the discovery of 2D graphene by Geim and Novoselov in 2004). Reaching movements are not left behind as they also have their own modeling research field, discussed in this section. First, we briefly discuss a few models that were or are used to model reaching movements and evaluate them in light of experimental data. Then, we detail the Optimal Feedback Control model which is, to date, the most popular model used to investigate reaching movements. We then present a functional neuroanatomy of this Optimal Feedback Control framework based on neural evidence and experiments performed on patients suffering from

diverse brain lesions. Finally, the limitations of the Optimal Feedback Control model, as known before the work presented in this thesis, are discussed and existing solutions presented.

2.1 Candidate models

Selecting the most appropriate model to investigate and simulate a physical or biological phenomenon is a tedious task because of the large number of possible models and the potentially infinite number of model parameters. Ideally, the selected model should be able to reproduce and explain if not most, many of the experimental features of the phenomenon. For reaching movements, the ideal candidate would be a model that explain the behaviors detailed in section 1. Therefore, the model candidate should be able to model the *baseline* behavior that integrates both sensorimotor (position and structure of the target for instance) and cognitive factors (target reward) but also a *feedback* component able to integrate the same factors in the feedback responses to perturbations. In the subsequent sections, we present and discuss some candidate models that were proposed to study upper-limb reaching movements.

2.1.1 Minimum-jerk model

Different control models were proposed for reaching movements through the years. In 1985, Flash and Hogan [Flash and Hogan, 1985] reported that the curvature of trajectories during reaching movements could be modeled using a *minimum-jerk model* that computes the movement trajectory from the minimum of the derivative of the acceleration. Mathematically, this model minimizes the following cost-function

$$C = \frac{1}{2}\int_0^{t_f} \left(\left(\frac{d^3x(t)}{dt^3} \right)^2 + \left(\frac{d^3y(t)}{dt^3} \right)^2 \right) dt, \tag{I.2}$$

where $x(t)$ and $y(t)$ are the time varying end-point coordinates of the simulated limb. This derivation, combined with the assumption that the velocity and acceleration are zero at the end and at the beginning of movement gives the following time-dependent expressions[9] for the x- and y-coordinates of the limb end-point

[9] These expressions can be obtained by minimizing equation I.2 for both coordinates

$$x(t) = x_0 + (x_0 - x_f)\left(15t^4 - 6t^5 - 10t^3\right), \tag{I.3}$$

$$y(t) = y_0 + (y_0 - y_f)\left(15t^4 - 6t^5 - 10t^3\right), \tag{I.4}$$

where x_0, y_0 and x_f, y_f are the initial and final positions, respectively. Even though this minimum-jerk model can reproduce the behavior observed as participants reached to a single target, it cannot model the online corrections to visual or mechanical perturbations, which makes it a bit too narrow to be adopted as a good model candidate for sensorimotor control. Moreover, this model does not contain any control component in that the agent cannot interact with the system. This model only explains the hand trajectories during movement but does not explain their causes.

2.1.2 Minimum-variance model

Similarly, in 1998, Harris and Wolpert [Harris and Wolpert, 1998] suggested that saccadic eye movements and reaching movements could be modeled using a *minimum-variance model* exploiting signal-dependent noise inherent to these movements. This model was able to explain the Fitts' law [Fitts, 1954] stating that there is power law between movement time and the target width. To model the reaching behavior, they modeled the limb dynamics using a discrete-time state-update equation

$$\mathbf{x}[t] = \mathbf{A}\mathbf{x}[t-1] + \mathbf{B}(\mathbf{u}[t] + \omega[t]) \tag{I.5}$$

where $\mathbf{x}[t] \in \mathbb{R}^n$ is the state of the system, $\mathbf{u}[t] \in \mathbb{R}^p$ is the motor commands, $\omega[t] \in \mathbb{R}^p$ is additive Gaussian noise and $\mathbf{A} \in \mathbb{R}^{nxn}$ and $\mathbf{B} \in \mathbb{R}^{nxp}$ are non-zero matrices. The optimal set of motor command is obtained by minimizing the sum of the positional variance V_t at the end of movement for a predefined time window

$$\min \sum_{t=T}^{t=T+R} V_t \tag{I.6}$$

where T is the prespecified movement duration and R is the time during which the hand has to stay in the target. Even though this model could reproduce experimental behaviors in the unperturbed baseline condition, it was unable to explain the feedback mechanism that are recruited in response to perturbations.

A series of models were introduced later in the field to integrate both the *baseline behavior* and the *feedback* mechanisms : the *impedance control model* [Burdet et al., 2001] and the *Optimal Feedback Control* model [Todorov and Jordan, 2002]. These will be developed further in the next sections.

2.1.3 Impedance control model

Impedance control posits that the agent compensates for perturbations by learning to stabilize unstable dynamics through selective control of impedance geometry. Concretely, this means that a modulation of joint stiffness is sufficient to reject any disturbances that would move the state of the system away from its nominal trajectory. For the mathematical derivation (see [Gomi and Kawato, 1996, Gomi and Kawato, 1997]) of this model of upper limb motor control, one has to start from the planar plant dynamics of the human arm using a second-order nonlinear differential equation

$$\psi\left(\ddot{\mathbf{q}}, \mathbf{q}, \dot{\mathbf{q}}\right) = \tau_{in}\left(\dot{\mathbf{q}}, \mathbf{q}, \mathbf{u}\right) + \tau_{ext} \tag{I.7}$$

where $\psi\left(\right)$ represents the dynamics of a two links arm and $\mathbf{q}$, $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ are the vector representing the angular position, velocity and acceleration respectively. The term τ_{ext} refers to the external torques applied to the arm (e.g. pertubations) and τ_{in} is the torques generated by the agent, with $\mathbf{u}$ being the motor command descending from the supraspinal nervous system. The left-hand side of equation I.7 can be developed as follows if one assume the arm to be a rigid body characterized by inertia and coriolis forces

$$\psi\left(\ddot{\mathbf{q}}, \mathbf{q}, \dot{\mathbf{q}}\right) = \mathbf{I}\left(\mathbf{q}\right)\ddot{\mathbf{q}} + \mathbf{H}\left(\dot{\mathbf{q}}, \mathbf{q}\right) \tag{I.8}$$

Using the variational principle, the motor command required to bring the arm from an initial state $[\mathbf{q},\ \dot{\mathbf{q}},\ \ddot{\mathbf{q}}]$ to a target state $[\mathbf{q}^*,\ \dot{\mathbf{q}}^*,\ \ddot{\mathbf{q}}^*]$ can be written

$$\tau = \mathbf{K}\left(\mathbf{q}^* - \mathbf{q}\right) + \mathbf{D}\left(\dot{\mathbf{q}}^* - \dot{\mathbf{q}}\right) + \mathbf{I}(\mathbf{q})\ddot{\mathbf{q}}^* + \mathbf{H}\left(\mathbf{q}, \dot{\mathbf{q}}\right) \tag{I.9}$$

The closed loop expression writes

$$\tau_{ext} = \mathbf{K}\left(\mathbf{q}^* - \mathbf{q}\right) + \mathbf{D}\left(\dot{\mathbf{q}}^* - \dot{\mathbf{q}}\right) + \mathbf{I}(\mathbf{q})\left(\ddot{\mathbf{q}}^* - \ddot{\mathbf{q}}\right) \tag{I.10}$$

With this closed loop expression, we can observe that the system behaves as a spring mass damper with $\mathbf{K}$ and $\mathbf{D}$ the spring and damping constant, respectively. Importantly, this model of upper limb control assumes that any deviation from the nominal trajectory is corrected by leveraging the viscoelasticity properties of the muscles. This relies on the assumption that during movement, the agent tries to follow a nominal trajectory which implies a feedforward planning of movement and that the impedance properties of the limb are sufficient to reject any disturbances to the system. This last point was questioned in a study thta demonstrated that joint stiffening would only have an impact for a small range of perturbation intensities [Crevecoeur and Scott, 2014]. Moreover, joint stiffening is theoretically unable to contain sustained mechanical perturbations as it is only comparing angle positions and velocities. Another work demonstrated that the motor system can behave like a trajectory controller only if the goal of the task is to follow a desired control strategy [Cluff and Scott, 2015], restricting even more the applicability of impedance control to human reaching movements.

2.1.4 PID controller

The motor command derived for the impedance control model (equation I.9) is somehow reminiscent of the PID controller [Ziegler and Nichols, 1942, Astrom and Hagglund, 1995]. Indeed, in a PID controller, the motor command is defined based on the error : $e(t) = r(t) - y(t)$ where $r(t)$ is the desired setpoint and $y(t)$ is the measured state of the system. The motor command of this controller is given by

$$\mathbf{u}\,(t) = K_p e(t) + K_i \int_{\tau=0}^{t} e(t)d\tau + K_d \dot{e}\,(t) \tag{I.11}$$

where K_p, K_i and K_d are the **P**roportional, **I**ntegral and **D**erivative real-valued constants of the PID controller. The desired behavior is obtained through the tuning of those **P**, **I**, and **D** parameters which is relatively straightforward for univariate systems [Astrom and Hagglund, 1995, Astrom and Hagglund, 1984, McMillan, 1983] and more intricated for multivariate systems [Wang et al., 2008, Wang et al., 1997, Koivo and Tantuu, 1991] because of potential interactions between multiple variables.

We can observe some similarity between equations I.9 et I.11 as both of them defined the motor command as a linear combination of derivatives or integrals of the error[10]. Importantly, defining the motor command this way does not provide any guarantee about the optimality of the control and its stability. Moreover, some tuning of the different constant terms is required in order to achieve a desired behavior [Astrom and Hagglund, 1995]. In the framework of impedance control for sensorimotor control, it is assumed that these gains are tuned because the central nervous system has learned the optimal impedance of its limbs [Burdet et al., 2001]. To model reaching behavior, these models will require to continuously modify the setpoint along a prespecified trajectory (that can for instance be obtained by the minimum jerk model). This assumes that the central nervous system performs *trajectory control*.

A strong argument against any sensorimotor control model involving one or several setpoints (e.g. the *minimum-jerk model*, the *impedance model* or *PID controller*) is that the behavior observed experimentally rejects this hypothesis. Indeed, when they move, humans and other animals exquisitely exploit the many degree of freedoms inherent to their limb (see section 1.2.1). The failure of these control models to explain reaching movements gave rise to the use of goal-directed control schemes to understand these movements.

In this thesis, we mainly consider the Optimal Control Framework as a theoretical canvas to model reaching movements. The rationale of this framework as well as the corresponding mathematical derivations are developed in section 2.2.

2.2 Optimal Feedback Control framework

The theoretical framework of optimal control has been translated from the field of control theory to the study of human motor control by Emo Todorov [Todorov and Jordan, 2002]. The basic premise of this theory is that the controller generates the command based on an inverse model of the system dynamics and guarantees convergence to the goal thanks to state feedback policies.

[10] To improve the parallelism, the error can be compared to the term $(\mathbf{q} - \mathbf{q}^*)$.

The idea underlying Optimal Feedback Control is that at any time, the central nervous system commands the limb with the motor command that is optimal to steer the system towards the desired state. This optimal command depends on the state of the system at that time and on the, supposedly known, dynamics of system. Importantly, this control model does not plan the trajectory ahead of movements because the motor command is defined as linear combination of the state vector, gathering all the latent variables of the controlled system.

2.2.1 Mathematical derivation

The optimal control problem aims at finding the best set of commands to apply to a system in order to achieve the goal defines by some criterion [Kirk, 2004]. Let us consider a dynamical system characterized unequivocally by the state vector $\mathbf{x}(t) \in \mathbb{R}^n$ (containing all the information relevant to the system) and the command vector $\mathbf{u}(t) \in \mathbb{R}^p$ (that corresponds to the action of the agent on the system). The system dynamics can be captured by the following differential equation:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{u}(t), t), \tag{I.12}$$

For the general optimal control problem, the agent's performance can be quantified thanks to a performance index captured by a *cost-function*. This performance index weights motor error and motor cost and can be expressed, without loss of generality as follows:

$$\mathbf{J}(\mathbf{x}, \mathbf{u}) = \underbrace{h(\mathbf{x}(t_f), t_f)}_{\text{motor cost}} + \underbrace{\int_{t_0}^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t)\, dt}_{\text{motor error}}, \tag{I.13}$$

where h and g are scalar functions and t_0 and t_f correspond to the initial time and to the final time that can be defined as a free parameter. In this framework, the set of optimal commands are the commands $\mathbf{u}^*(t)\ \forall t \in [0, t_f]$ which minimizes equation (I.13) for the system governed by equation (I.12).

2.2.2 The LQG controller

In this work, we mainly consider a specific case of optimal control, the Linear Quadratic Regulator, where the system dynamics is assumed to be *linear* and the cost function is assumed to be *quadratic* with respect to the state. In this specific case, the discrete-time[11] state dynamics can be defined as

$$\mathbf{x}\left[t+1\right] = \mathbf{A}\mathbf{x}\left[t\right] + \mathbf{B}\mathbf{u}\left[t\right] + \boldsymbol{\xi}\left[t\right], \tag{I.14}$$

where $\mathbf{A}$ and $\mathbf{B}$ are real-valued matrices and $\boldsymbol{\xi}\left[t\right] \sim \mathcal{N}\left(\mathbf{0}^{n\text{x}1}, \boldsymbol{\Sigma}_{\text{motor}}\right)$ is additive gaussian white motor noise whose covariance matrix is $\boldsymbol{\Sigma}_{\text{motor}}$.

The cost-function is also refurbished and becomes a quadratic function of state and command that can be expressed as

$$\mathbf{J}\left(\mathbf{x}, \mathbf{u}\right) = \sum_{i=1}^{N} \left\{ \mathbf{x}\left[i\right]^{T} \mathbf{Q}\left[i\right] \mathbf{x}\left[i\right] + \mathbf{u}\left[i\right]^{T} \mathbf{R}\mathbf{u}[i] \right\} \tag{I.15}$$

where N is the fixed time horizon, $\mathbf{Q}\left[i\right] \geq 0 \ \ \forall i$ and $\mathbf{R} > 0$ are respectively the matrices that penalize motor errors at any time step[12] and the motor commands. This cost function allows to translate the high-levels goals (ie. reach for an object) into the real-time sensorimotor control policies used to achieve these goals [Todorov, 2004].

In biological systems such as the upper limb explored in this thesis, the state of the system is not fully observable because of noise and delays in the sensory feedbacks. For this reason, it is needed to introduce a new quantity $\mathbf{y} \in \mathbb{R}^{m}$ the observation which is defined by:

$$\mathbf{y}\left[t\right] = \mathbf{H}\mathbf{x}\left[t\right] + \boldsymbol{\omega}\left[t\right], \tag{I.16}$$

where $\mathbf{H}$ is the observation matrix (often set to $\mathbf{1}^{m\text{x}m}$) and $\boldsymbol{\omega}\left[t\right] \sim \mathcal{N}\left(\mathbf{0}^{m\text{x}1}, \boldsymbol{\Sigma}_{\text{sensory}}\right)$ is additive gaussian white sensory noise whose covariance matrix is $\boldsymbol{\Sigma}_{\text{sensory}}$.

[11] The discrete-time state space representation is obtained using the Euler integration method on the continuous-time state space representation

[12] Traditionally, only the very last or the few last time steps have non-zeros $\mathbf{Q}$ matrices as only the goal target matters.

Since the state if not fully observable because of delays and noise, the state of the system has to be estimated by optimally combining sensory inputs and internal prediction based on a copy of the motor commands. This optimal combination of sensory evidence and internal prediction can be summarized as follows:

$$\hat{\mathbf{x}}[t] = \mathbf{A}\hat{\mathbf{x}}[t] + \mathbf{B}\mathbf{u}[t] + \mathbf{K}[t](\mathbf{y}[t] - \mathbf{H}\hat{\mathbf{x}}[t]), \tag{I.17}$$

$$\mathbf{K}[t] = \mathbf{A}\boldsymbol{\Sigma}[t]\mathbf{H}\left(\mathbf{H}\boldsymbol{\Sigma}[t]\mathbf{H}^T + \boldsymbol{\Sigma}_{\text{sensory}}\right)^{-1}, \tag{I.18}$$

$$\boldsymbol{\Sigma}[\boldsymbol{t}+\mathbf{1}] = \boldsymbol{\Sigma}_{\text{motor}} + (\mathbf{A} - \mathbf{K}[t]\mathbf{H})\boldsymbol{\Sigma}[t]\mathbf{A}^T, \tag{I.19}$$

where $\mathbf{K}[t]$ are the *Kalman gains* whose role are to weight sensory evidence and internal prediction in state estimation and $\boldsymbol{\Sigma}[t]$ is the covariance matrix of the state estimate $\hat{\mathbf{x}}[t]$ at each time step.

In this model, the optimal command at any time is proportional to the estimated state of the system $\hat{\mathbf{x}}[t]$:

$$\mathbf{u}^*[t] := -\mathbf{L}[t]\hat{\mathbf{x}}[t], \tag{I.20}$$

where $\mathbf{L}[t] \in \mathbb{R}^{pxn}$ are the *control gains* matrices computed, in the finite horizon case that we developed here, recursively as follows :

$$\mathbf{L}[t] = \left(\mathbf{R} + \mathbf{B}^T\mathbf{S}[t]\mathbf{B}\right)^{-1}\left(\mathbf{B}^T\mathbf{S}[t]\mathbf{A}\right), \tag{I.21}$$

$$\mathbf{S}[t-1] = \mathbf{Q}[t] + \mathbf{A}^T\mathbf{S}[t](\mathbf{A} - \mathbf{B}\mathbf{L}[t]), \tag{I.22}$$

$$\mathbf{S}[N] = \mathbf{Q}_f, \tag{I.23}$$

where $\mathbf{Q}_f$ is the cost matrix defining the end point penalty on motor error and $\mathbf{S}[t]$ is a proxy that allows to compute the cost-to-go from the current state of the system.

This LQG[13] controller is a good candidate to model reaching movements as it reproduces several experimental features (see section 2.2.3) and, most importantly, includes both baseline and feedback mechanisms within the same model. The specific application of this controller to reaching movements is discussed in section 2.2.3.

[13]LQG stands for **L**inear state dynamics, **Q**uadratic cost function and **G**aussian additive noise

2.2.3 OFC applied to reaching movements

The model formulation developed in section 2.2.2 has been used, with some variation, in many studies to investigate the ability of this framework to reproduce experimental results collected during laboratory tasks.

Figure I.4 represents a schematic block diagram of the optimal feedback control framework in the version that is the most commonly used to model human reaching movements. During the first step, *movement planning*, the agent collects information about the task and environment to select the appropriate control policy, $\pi\left(\boldsymbol{x}\left[t\right], t\right)$, which is afterwards transferred to the *controller*. The goal of the controller is to continuously monitor the *state estimate* during movement execution in order to provide the appropriate *motor command*. *Proprioceptive* and *sensory feedbacks* are combined with the internal prediction based on the efference copy to optimally estimate the state of the system.

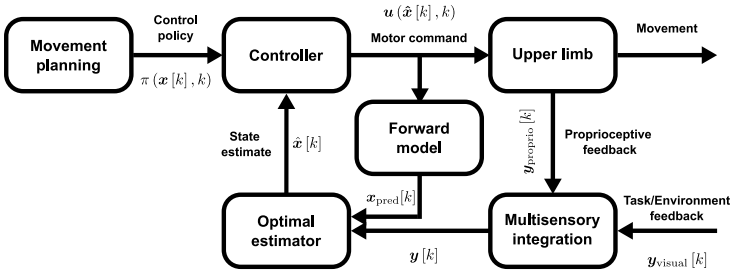

Figure I.4: Schematic block diagram of the optimal feedback control framework. The control policy selected during movement planning is used to execute movements. At any time step, the optimal motor command is sent to the plant (i.e. the upper limb) which results in movement. A copy of this motor command (i.e. efference copy) is sent to the optimal estimator that combines it with sensory feedback about the limb and the task to estimate the state of the system. This estimated state of the system is then fed back to the controller.

A hallmark of Optimal Feedback Control is the so-called *minimum intervention principle* stating that the controller does not correct noise and disturbances if they do not interfere with the success of the task [Todorov and Jordan, 2002, Liu and Todorov, 2007]. This means that noise and disturbances could accumulate in task-irrelevant directions, which experimentally corresponds to wider end-point distributions when reaching to wider targets [Knill et al., 2011, Nashed et al., 2012, Lowrey et al., 2017]. In addition to this minimum intervention principle,

OFC also reproduces previous findings such as the bell-shaped velocity profiles of reaching movements [Morasso, 1981, Flash and Hogan, 1985]. These first results demonstrated that OFC was a good candidate to model the baseline behavior of motor control as demonstrated by its ability to model motor variability and the characteristics behaviors observed during experiments [Diedrichsen et al., 2010].

Besides these fundamental properties of reaching movement, OFC is also able to model the behavior observed in presence of various disturbances. Indeed, it has been demonstrated that such models could predict participants' behavior as they reached to a narrow or wide target in a study where mechanical perturbations could be applied to participants' hand [Nashed et al., 2014]. In this study, the model predicted that participants will correct the mechanical perturbations only if the target was narrow, which is what their and others' experimental results confirmed [Knill et al., 2011, Lowrey et al., 2017]. Another study investigated whether this model could explain the behavior observed when participants reached to target whose true location was uncertain[14][Izawa and Shadmehr, 2008]. They reported that, similarly to their participants' behavior, the model predicted that the motor responses to perturbations depended on the target uncertainty which confirmed an implication of target information in the feedback loop. A third study investigated the ability of OFC to model bimanual reaching movements [Diedrichsen, 2007]. They observed that whether participants had to control two cursors (one with each hand) or only one cursor (located at the average position of both hands) changed the behavior as they could observe changes in the trajectories of both hands only in the one cursor condition. This demonstrated that bimanual control is optimally adjusted to the current task requirements.

These results confirmed that Optimal Feedback Control is a good candidate to model reaching movements as it explains both baseline behavior and feedback mechanisms. Furthermore, it integrates low-level sensorimotor or high-level cognitive factors in the control policy.

[14]In this experiment, Izawa and colleagues modulated the target blur to add uncertainty in the target location.

2.3 Neural correlates of reaching movements

An outstanding question is whether the central nervous system can actually support optimal feedback control and, if yes, how this control policy is implemented in the brain. These questions were addressed in two reviews that aimed at providing a neuroanatomy of motor control [Scott, 2016, Shadmehr and Krakauer, 2008]. A schematic representation of this neuroanatomy is represented in Figure I.5 where the different functions have been associated to the different brain regions.

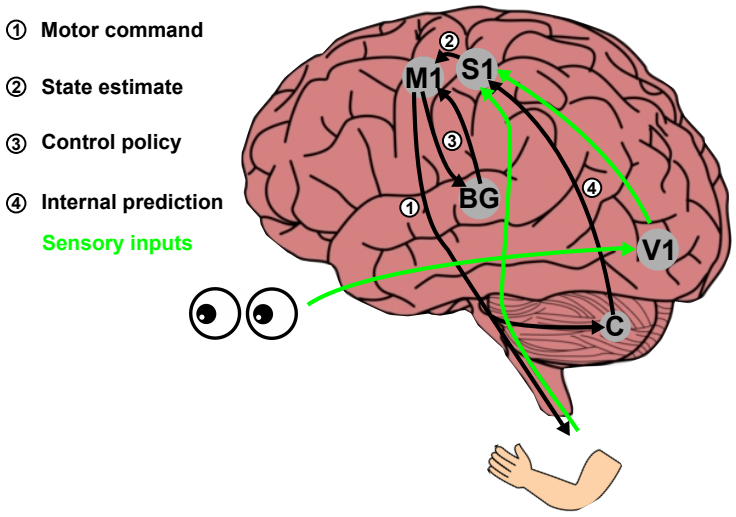

Figure I.5: Schematic representation of the neural implementation of optimal feedback control in the human brain. The task goal and control policy **3** are defined in the basal ganglia (**BG**) and afterwards used by the primary motor cortex (**M1**) before and during movement. The motor command **1** sent by the primary motor cortex to the limbs depends on the state estimation **2** performed by the sensory cortex (**S1**). The sensory cortex optimally combines sensory inputs (green) and the internal prediction **4** performed by the cerebellum (**C**).

2.3.1 Basal ganglia

Shadmehr and colleagues suggested that the *basal ganglia* are responsible to integrate the expected costs and task reward into a cost function used by the feedback controller [Shadmehr and Krakauer, 2008]. Indeed, Mazzoni and colleagues [Mazzoni et al., 2007] demonstrated that Parkinson's Disease (PD) patients were more sensitive to movement cost. They suggested that these behaviors could link the role

of basal ganglia in motor control to the motivation-reward framework in which basal ganglia play a central role [Schultz, 2006]. Cells within the basal ganglia have indeed been shown to fire more before rewarding movements [Kawagoe et al., 1998]. Moreover, the occurrence of a stimulus associated with more reward involved a burst of dopamine [Matsumoto and Hikosaka, 2007] followed by more vigorous movements [Tachibana and Hikosaka, 2012]. These results and many others demonstrated that the basal ganglia are not only involved in reward-based reinforcement learning [Schultz, 2006] but also plays a primordial role during movement planning and control policy selection.

Consistently with the hypothesis that the basal ganglia integrate the expected costs and task rewards, this structure is thought to influence movement vigor. Indeed, implicit motivation associated with the task and the control of saccades vigor are supported by the same system [Reppert et al., 2015, Xu-Wilson et al., 2009]. A deficit in movement vigor can be observed in PD patients when they can freely choose their moving speed [Baraduc et al., 2013, Mazzoni et al., 2007]. This deficit can be recovered partially through deep brain stimulation [Baraduc et al., 2013]. These results therefore suggest the implication of the basal ganglia in the modulation of movement vigor ([Dudman and Krakauer, 2016, Turner and Desmurget, 2010] for reviews). Interestingly, the basal ganglia also control the urgency associated with motor decisions [Thura and Cisek, 2017] which is a factor that is also known to impact the control strategies underlying reaching movements [Crevecoeur et al., 2013].

2.3.2 Premotor and motor cortices

Premotor and motor cortices have been proposed as the brain areas responsible for the continuous application of the feedback control policy [Kalaska et al., 1997, Scott, 2012, Shadmehr and Krakauer, 2008]. Perturbation studies using transcranial magnetic stimulation (TMS) targeting these areas demonstrated that the feedback responses to perturbations were influenced by these pulses suggesting a central role of these areas in the feedback control policy [Shemmell et al., 2009, Kimura et al., 2006, Pruszynski et al., 2011]. Direct recording of individual [Churchland and Shenoy, 2007, Herter et al., 2006, Herter et al., 2009, Kalaska, 2009] and population of neurons [Churchland et al., 2012, Kaufman et al., 2014,

Shenoy et al., 2013] in these cortical areas revealed a tuning of the individual activities to the target position or the perturbation as well as population level dynamics correlated to behavior. This confirmed the hypothesis that the hub for the application of the control policy during movement execution, but also during movement preparation [Churchland et al., 2012, Kaufman et al., 2014], is located in these cortex areas.

2.3.3 Cerebellum

In this Optimal Feedback Control framework, the role of the *cerebellum* is to predict the sensory consequence of the motor commands through internal models of the system dynamics [Wolpert et al., 1998, McNamee and Wolpert, 2019]. An internal model can be defined as a predictive mechanism that exploits the agent's knowledge about the system dynamics to predict the next state of the system based on the current state and the motor commands. The role of the cerebellum in this process has been revealed thanks to studies on patients with cerebellar damage. It has been demonstrated that these patients execute movements similar to control ones but with over- or under-shooting of the target and oscillations when reaching [Bastian et al., 1996, Therrien et al., 2016] or when pinching objects [Muller and Dichgans, 1994]. The central role of cerebellum in sensory prediction [Ebner and Pasalar, 2008, Brooks et al., 2015] is therefore a central element in Optimal Feedback Control as it is the neural substrate that conveys the predictive state of the system primordial to the state estimation.

2.3.4 Parietal cortex

The last process of OFC that has to be discussed is the optimal state estimation that combines sensory information with internal prediction. A neural structure usually associated with this process is the *parietal cortex* (PPC) [Shadmehr and Krakauer, 2008]. Multiple studies on patient suffering from lesions to the posterior parietal cortex demonstrated that this is important for spatial perception, association of sensory signals, directing attention, visuomotor control and movement planning [Wolpert et al., 1998, Rushworth et al., 1997, Gréa et al., 2002]. A perturbation study using TMS pulses targeted at PPC demonstrated its crucial role in the feedback responses to visual perturbations

[Desmurget et al., 1999]. Indeed, Desmurget and colleagues suggested that the parietal cortex plays a central role in the ability of the motor system to estimate the future state of the limb. This was confirmed in a study that revealed a activation of the parietal lobe (revealed by fmri imaging) in presence of target jumps suggesting that this area was recruited in the state feedback loop used to correct that perturbation [Chapman et al., 2002]. These results probed the development of the modeling of the role of PPC in the forward modeling, central to feedback control in fast reaching movements [Desmurget and Grafton, 2000].

This non-exhaustive atlas of the neuroanatomy of Optimal Feedback Control demonstrates that this is a distributed process whose various parts are handled at brain areas. This distributed characteristic is extremely interesting as it allows for the separate investigation of the different part of the controller to, for instance, investigate the impact of various parameters on the different elements of this process. Moreover, these different physiological evidences support the hypothesis that Optimal Feedback Control is a good candidate to model human reaching movements.

2.4 Beyond Optimal Feedback Control

Optimal Feedback Control, as presented above, appears to be a serious candidate to model reaching movements. First, this theoretical framework is able to capture several experimental results. From the movement variability, captured by the minimum intervention principle, to the exquisite and complex responses to mechanical and visual perturbations, handled through feedback loops, many experimental results can indeed be reproduced by this model. Second, Optimal Feedback Control possess a plausible neural implementation of fine motor control as it could use the different neural structures for the purpose they have experimentally be associated with.

However, this theoretical framework, as introduced by Todorov [Todorov and Jordan, 2002], possess some caveats since there are still some phenomena that it cannot explain. The goal of this section is to briefly discuss a non-exhaustive list of areas where the aforementioned implementation of OFC lacks some depth to explain the various facets of the experimental results. In this section, we first discuss the, somehow, limiting hypotheses underlying optimal feedback control and more

specifically the LQG controller introduced in section 2.2.2. The second part of this section covers the problem of time horizon, corresponding to the pre-planned movement time, used in the LQG formulation. The next section is somehow connected to the previous one as it investigates the fact that the control policy is decided prior to movement onset and that all the feedback gains are fixed within a movement. The last part of this section covers some parameters and factors that experimentally demonstrated an impact on movement control while not being easily integrable within the OFC framework.

2.4.1 The LQG hypotheses

The Optimal Feedback Control, developed in section 2.2.2 is implemented for reaching movements using the Linear Quadratic Gaussian controller. This controller assumes that the plant dynamics (equation I.12) is linear with respect to the state of the system $\mathbf{x}(t)$. It also assumes that the performance index, captured by the cost function (equation I.13), is quadratic with respect to the state $\mathbf{x}$ and motor commands $\mathbf{u}$. Finally, it assumes that additive Gaussian white noise corrupts output measurements. These three hypotheses are somewhat restrictive and we briefly discuss each of them here.

Limb dynamics is highly non-linear [Delp et al., 2007, Schaffelhofer et al., 2015] and approximating it with a point-mass model as we did in section 2.2.2 might lead to some gross approximation errors. The LQG method alone does not allow to tackle these non-linear dynamics but Li and Todorov developed an expansion of this method that could accommodate these dynamics [Li and Todorov, 2004, Todorov and Li, 2005]: the ILQG method. Briefly, this method uses iterative linearizations of the non-linear dynamics around the nominal trajectory. This nominal trajectory is then improved thanks to modified Riccati equations [Todorov and Li, 2005]. This method significantly improves the behavior of the LQG on non-linear dynamics and allows to model feedback control policy. However it is still based on first order approximation of the system dynamics and it does not have anything regarding constraints on the state or on the commands.

The quadratic cost-function is also a limiting hypothesis as it restrains the class of problems that can be modeled using LQG. Indeed, any cost function that cannot be expressed as a quadratic function of the state or that cannot decently be approximated by such a function requires subtle selection of the cost-function to be accommodated. For example, to model obstacles in the environment, one has to use via points to restrain the trajectories around them [Nashed et al., 2012] since a correct implementation of the obstacles would require highly non-quadratic cost function.

Finally, additive Gaussian noise is not the only source of noise in the motor systems as it has been demonstrated that signal dependent noise[15] is also a non negligible parameter of the upper limb dynamics [Harris and Wolpert, 1998]. This problem can however be tackled in the optimal control framework thanks to methods that integrate the knowledge of this signal dependent noise in the optimal state estimation [Todorov, 2005, Crevecoeur et al., 2011].

2.4.2 Finite time horizon

In the backward recursion presented in equations I.21-I.23, the control gains were computed recursively starting from the last time step corresponding to $t = N$ where N is a predefined number of time steps. Predefining this number of time steps is equivalent to predefining the time-horizon[16] of the controller. This means that the agent is supposed to know movement duration before initiating movement which is very unlikely given uncertainties about the occurrence of perturbations that could impact movement duration and the variability in movement duration inherent to reaching movements.

A solution to tackle this problem is to use infinite horizon optimal controller [Phillis, 1985]. Briefly, the infinite horizon controller differs from the finite horizon formulation in that it does not approximate the feedback and estimator gains as the LQG does but instead computes and applies the steady state solution throughout the whole episode [Phillis, 1985, Qian et al., 2013]. This alternative formulation of the LQG is able to model reaching movements and to grasp some

[15] Signal dependent noise is additive noise proportional to either the state of the system $\mathbf{x}$ or to the motor command $\mathbf{u}$.

[16] The time-horizon is defined as the number of time steps in the future for which the controller is computing the motor commands

of their properties such as the Fitt's law and speed accuracy trade-off [Qian et al., 2013]. It has also been suggested that this model could explain the behaviors observed during target and cursor jumps paradigm [Dimitriou et al., 2013] could also be explained by an infinite horizon formulation of this model as demonstrated in the paper by Li and colleagues [Li et al., 2018].

This infinite horizon formulation means that movement duration is not a predefined parameter of the control policy but rather would be a consequence of this control policy and could depend on various factors such as target structure [Nashed et al., 2012, Lowrey et al., 2017] or target reward [Shadmehr et al., 2010a, Haith et al., 2012] for instance.

2.4.3 Preplanned control for each state

Congruent with the discussion about the time horizon (see section 2.4.2), the OFC framework presented above assumes that the control policy is immutable once it has been selected. This means that if a task-related parameter is modified during movement in such a way that the selected cost-function is not suitable anymore, the adopted control policy will be maintained and the movement probably failed.

This problem was first introduced by Dimitriou and colleagues [Dimitriou et al., 2013] that investigated participants' behavior as they were reaching to a target that could change location during movements as well as the hand-aligned cursor. They observed a modulation of the feedback gains during movements for different perturbation onsets that they explained using model predictive control (see [Lee, 2011] for review). Briefly this framework computes, at every time step, the control policy for a prespecified time horizon N_{hor} and only applies the very first command, others being discarded.

Model predictive control formulation of sensorimotor control has been recently explored to explain the results observed by Dimitriou and colleagues [Dimitriou et al., 2013] in a study where the control policy was recomputed whenever a change in target or cursor locations occurred [Cesonis and Franklin, 2020]. However, it has been demonstrated that these results could also be explained by state feedback control while using a infinite-horizon controller [Li et al., 2018]. In a recent study [Cesonis and Franklin, 2021], it has been suggested that infinite horizon controllers are however not enough to explain this behavior as they

could not explain the modulation of feedback gains during movement. They suggested that model predictive control is a better candidate for these results and that the model predictive mechanism is triggered whenever a perturbation (visual or mechanical) occurs.

2.4.4 Missing factors

Besides the different points inherent to the LQG formulation that we discussed above, optimal control fails at modeling some specific task paradigms that are widely used in the field of sensorimotor control. Indeed, since the control policy is computed backward starting from the goal target, it might be complex to properly grasp things such as target redundancy or multiple targets. To model targets with different redundancy along different axes, Nashed and colleagues [Nashed et al., 2012] modified the parameters of the cost function associated with the distance aligned with the larger redundancy. In another study [Nashed et al., 2014], they tackled the problem of multiple targets by comparing the cost-to-go functions to the different targets after perturbation to determine which target will be reached after perturbation. However, it is still unclear how parameters such as target reward or multiple targets are integrated into the selection of the control policy.

3 Aim of this work

Despite being an active research topic for decades, human sensorimotor control and more specifically the study of upper limb reaching movements still has many unanswered questions. The understanding of the control policy used by humans for sensorimotor control and its neural implementation is a promising field of research as it could help understand the implication on some neural impairments and eventually suggest future therapies.

The goal of the work presented in this thesis is to uncover some mechanisms underlying reaching control by the mean of experimental paradigms. We mainly focus on the points presented in sections 2.4.3 and 2.4.4 [17], and investigate whether the parameters known to influence behavior during movement planning and static decision-making

[17] Namely the possibility to change the control policy online and the integration of multiple alternative targets associated with different rewards

could also influence the control policies during movement execution. For this purpose, we present the different experiments performed as part of this thesis. The first two series of experiments investigate online adjustments in control policies while the last set of experiments reveals the delicate imbrication of robustness and flexibility in the online decision processes. As stated earlier in the introduction, the work presented in this thesis is at the border of experimental and modeling work. Therefore, most of the experimental results presented in the subsequent chapters are analyzed through the lens of control theory concepts. These control theory concepts were further explored in a final chapter where we developed qualitative control models that provides modeling support for the conclusion we drew from the experimental data.

In Chapter II, we investigate to what extent the control policies used during movement can be adjusted online. We performed a series of experiments combining online change in task demands (structure of the goal target or obstacles) and mechanical perturbations to study whether the control policies can be adjusted online to the new task requirement. We report consistent online adjustments in control policy, as reflected by change in feedback responses to mechanical perturbations, coherent with the change in task requirements within as little as 150ms following the visual perturbation onset. These results demonstrate that the control policies used by humans to perform reaching movements are not frozen during movement execution and can be altered while moving in response to change in task demands.

In Chapter III, we elaborate on the findings of Chapter 2 by investigating the nature of the processes responsible for the online adjustments in control policy. More specifically, we used different changes in target structure during movements to determine whether these online adjustments in control policy resulted from a switch before different prespecified motor plans or whether they could result from a continuous feedback mechanism adjusting the control policy. We report that online adjustments in control policy, as reflected by feedback responses to mechanical perturbations, differed for discrete and continuous changes in target structure. Moreover, we demonstrated that these online adjustements were tuned to the specific rate of change in task demands. These results demonstrate that these online adjustments in control

policy continuously capture dynamic factors associated with the task demands through a feedback mechanism that regulates the control policy.

In Chapter IV, we first investigate the effect of target reward on the control policy and reveal that increasing the target reward results in an increase of feedback gains. This increase of the feedback gains resulted in faster movements and larger responses to mechanical perturbations which could be attributable to the use of a more robust control policy. Hypothesizing that this increase of feedback gains would be detrimental to one's ability to change target online, we designed a second experiment where participants had to perform reaching movements to any of three targets. The reward associated with each target could differ and mechanical perturbations were used to elicit switching behavior. We observed that participants' flexibility was dependent on the feedback gains selected for the movement. A causal link between these feedback gains and the ability to switch target during movement was further revealed in a third experiment involving background forces used to artificially increase muscle activity. Altogether, these results demonstrated a competition between movement vigor (characterized by the increase in feedback gains) and movement flexibility (characterized by the ability to change target during movement).

In Chapter V, we extend the theoretical framework of Optimal Feedback Control with two different algorithms aiming at explaining the experimental findings of chapters II, III, and IV. The first algorithm proposes a dynamical adjustment of the control policy resulting in online modification of the feedback gains during movement. Using this implementation we reproduce the experimental results of chapter II and III. The goal of the second algorithm is to expand the control problem to composite non-convex cost-functions such as those corresponding to the experiments of chapter IV. This second algorithm was able to predict the outcome of online motor decisions by comparing the cost associated with the different outcomes. With these two algorithms, we propose a dynamical extension to the OFC framework that could model and explain the behaviors observed in the experiments detailed in the present thesis.

In Chapter VI, we discuss the results of Chapters II-V as well as their limitations and implications with respect to the Optimal Feedback Control framework. We also exposed some unanswered questions that this thesis has raised and suggest further studies that could be conducted to tackle these questions.

In all the behavioral and experimental results presented in this thesis manuscript support the hypothesis that movement planning and execution are intricated mechanisms. Indeed, by demonstrating that the factors that influence movement planning (e.g. the task demands or the reward associated with the task) also influence movement execution, we revealed that the border between movement planning and execution is thinner that previously suggested. The novelty of this work is that we studied the human sensorimotor control system through a control theory point of view which helped us develop novel experimental paradigms and address new research questions. We specifically manipulated non-static task-related cost parameters during movements to unveil dynamical planning of movements during their execution. Moreover, we came up with a proposition to update control models of the human sensorimotor control in order to explain the observed results. Together, we believe that the work presented here consists in a promising starting point for a better understanding of human motor control.

4 Publications and communications

Published

- **De Comite A., Crevecoeur F., and Lefèvre P.** Online modification of goal-directed control in human reaching movements. *J Neurophysiology* 125(5), 1883-1898, 2021

Submitted

- **De Comite A., Crevecoeur F., and Lefèvre P.** Reward-dependent selection of feedback gains impacts rapid motor decisions. *eNeuro*, first submitted 5th of October 2021

In preparation

- **De Comite A., Crevecoeur F., and Lefèvre P.** A feedback mechanism underlying online adjustments of goal-directed control in humans reaching movements.

Proceedings

- **De Comite A., Crevecoeur F., and Lefèvre P.** Modification of the control law in reaching movements induces by online target change. *48th Annual Meeting of the Society for Neuroscience*, San Diego, CA, 2018 [Poster]
- **De Comite A., Crevecoeur F., and Lefèvre P.** How does reward of the goal target modify the control law in human reaching movements? *49th Annual Meeting of the Society for Neuroscience*, Chicago, IL,2019 [Poster]
- **De Comite A., Crevecoeur F., and Lefèvre P.** Influence of target reward on reaching control and rapid motor decisions *Annual Meeting of the Neural Control of Movement Society*, Dubrovnik, 2020, *Cancelled because of COVID* [Poster]
- **De Comite A., Crevecoeur F., and Lefèvre P.** Influence of target reward on rapid reaching decisions, *Neuromatch Conference 3*, Online, 2020 [Talk]
- **De Comite A., Crevecoeur F., and Lefèvre P.** Online adjustments in control policy in human reaching movements reflect dynamical changes in target structure, *Annual Meeting of the Neural Control of Movement Society*, Online, 2021 [Poster]

Chapter II

Online modifications of control policy in human reaching movements

Comprehension follows perception

Philip K. Dick

Published as : De Comite A., Crevecoeur F., and Lefèvre P. Online modification of goal-directed control in human reaching movements, *Journal of Neurophysiology*, 125 (5), 1883-1898, 2021

Humans are able to perform sophisticated reaching movements in a myriad of contexts thanks to flexible control policies tuned to the task goal and to the environmental context. However, it remains unknown whether these control policies can be adjusted online, during movement. The goal of the study presented in this chapter was to determine whether the factors that regulated control policies during planning could also modify the execution of an ongoing movement following abrupt changes in task demand. More precisely, we investigated whether, and at which latency, feedback responses to perturbation loads followed the change in the target structure or in environment. We changed the target width (narrow square or wide rectangle) to alter

the task redundancy, or the presence of obstacles to induce different constraints on the reach path, and assessed based on surface EMG recordings when the change in visual display modified the feedback responses to mechanical perturbations. Task-related EMG responses were detected within 150ms of a change in target structure. Considering visuomotor delays of ~ 100ms, these results suggest that it takes 50ms to change control policy within a trial. An additional 30ms delay was observed when the change in context involved sudden appearance or disappearance of obstacles. Overall, our results demonstrate that the control policy within a reaching movement is not static: contextual factors which influence movement planning also influence movement execution at surprisingly short latencies. Moreover, the additional 30ms associated with obstacles suggest that these two types of changes may be mediated via distinct processes.

1 Introduction

Humans perform dozens of reaching movements every day in a broad range of contexts. Task parameters such as the size and orientation of the goal target or the presence of obstacles influence these movements [Chapman and Goodale, 2008, Glazebrook, 2018, Howard and Tipper, 1997, Knill et al., 2011, Lacquaniti et al., 1986, Nashed et al., 2012] . Indeed, these factors influence planning policies as reflected in the ability of both visual and mechanical perturbations to affect the kinematics of movements and the feedback control policies. For instance, when reaching for a wide rectangular target, participants do not correct deviations aligned with the wider dimension of that target as they do when reaching for a narrow square [Knill et al., 2011, Nashed et al., 2012]. Similarly, participants modify their reaching trajectories to navigate around obstacles dependent on inertial factor [Sabes et al., 1998] and following perturbations [Nashed et al., 2012]. Together, these results provide experimental evidence that control policies used by humans are tuned to the task goal and environmental context. These results also demonstrate that a continuous stream of sensory information regarding the task and its environment is conveyed to the neural feedback controller [Crevecoeur et al., 2020a, Izawa and Shadmehr, 2008, Scott, 2016].

Optimal Feedback Control (OFC) framework was used to understand experimental findings related to humans motor control. This theory suggests that the control policy used to perform movements is tuned to the task goal and to its environmental context. In this framework, task constraints and knowledge of the limb dynamics are used to derive an optimal goal-dependent control policy defining the feedback gains applied during movement [Liu and Todorov, 2007, Scott, 2004, Todorov and Jordan, 2002]. This goal-dependent control policy is proposed to produce both the voluntary motor commands toward the target and the corrective feedback responses to external perturbations. Considering that different cost-functions are related to different tasks, OFC theory has been able to reproduce many of the experimental features mentioned above. Importantly, in previous studies, the cost-functions used to characterize task-dependent control policies were typically varied across conditions, including for instance different targets in a tracking task or different target shapes [Gallivan et al., 2016b, Lowrey et al., 2017, Nashed et al., 2012, Nashed et al., 2014]. These studies have not considered changes in cost-function corresponding to changes in task demands occurring during a movement. However, under natural conditions, it is conceivable that control policies and cost parameters vary over time. For example, in field sports the obstacles (e.g., opponents) move quite a bit. To date, it remains unknown whether and how quickly the nervous system can update its control policy following a change in cost-function.

Dynamic updates in cost-function, followed by adjustments in control policy, have been described in theory in the framework of Model Predictive Control ([Lee, 2011] for review). The basic premise of this theory is that changes are addressed with successive re-computations of control policy. This framework was used by Dimitriou and colleagues [Dimitriou et al., 2013] to interpret how feedback gains temporally evolved in a reaching task when both the goal target and a hand-aligned cursor jumped during movements. These authors observed changes in feedback gains when the target jumped during movements suggesting an online re-computation of control policy evoked in response to jumping of the target. It remained unclear how quickly the control policy was re-computed, which is the main purpose of this chapter.

To address this issue, a specific paradigm was needed because a possible shortcoming of the approach based on target jumps is that it does not necessarily imply a re-computation of feedback gains [Li et al., 2018]. Instead, feedback responses in this case could result from state-feedback control associated with long (infinite) time horizon [Qian et al., 2013]. The change in target structure imposes a selective change only linked to the target redundancy. Therefore, we could expect changes in the feedback gains aligned with this target redundancy [Liu and Todorov, 2007]. In this manner, feedback responses to perturbations do not simply follow from a state-feedback correction but also involve an adjustment of the control policy. Specifically, this paradigm is different from the RESIST/LET GO paradigm in which a binary verbal instructions about how to respond to the perturbation are provided [Hammond, 1956, Rothwell et al., 1980, Shemmell et al., 2009]

To achieve this goal, we leveraged the impact of a change in target structure on reach control to investigate whether an online modification of goal evoked an adjustment of feedback gains. Specifically, the target was switched during movements from a narrow square to a wide rectangle or vice-versa. This switch represented a change from constraints imposed on the two coordinates of the endpoint to redundancy along the main axis. In this context, two alternative hypotheses exist: if the control policy remains fixed within each movement, we expect that a change in target shape will have no impact on feedback responses to perturbation loads, and no changes in movement kinematics should be detected. In contrast, if a change in task demands evokes changes in the control policy online, then we expect that the feedback responses to the mechanical loads and subsequent kinematics adjust to the new target. Consistent with the latter hypothesis, we found that participants produced corrective responses adjusted to the change in target shape as well as to changes in environmental context (e.g., the presence of obstacles). Strikingly, evoked changes in muscle activity were observed in as little as 150ms following the change in target shape, while we measured a slightly longer latency with obstacles. These results demonstrate that control policy used to perform reaching movement can be updated rapidly during movement execution.

2 Methods

2.1 Participants

Three distinct groups of participants were enrolled in this study. The first group performed Experiment 1 and included 14 right-handed participants (5 females) ranging in age from 18 to 29 years. The second group performed Experiment 2 and included 17 right-handed participants (7 females) ranging in age from 19 to 25 years. The last group performed Experiment 3 and included 8 right-handed participants (4 females) ranging in age from 19 to 26 years. All groups of participants were naïve to the purpose of the experiments and had no known neurological disorder. The local ethics committee of the Université catholique de Louvain approved the experimental procedure and participants provided their written informed consent.

2.2 Setup

Participants sat comfortably on an adjustable chair in front of a Kinarm end-point robotic device (Kinarm Technologies, Kingston, ON, Canada). They were asked to grasp the handle of the right robotic arm with their right hand. The robotic arm allowed movements in the horizontal plane. Direct vision of both the participant's hand and the robotic arm was blocked by a semi-transparent mirror placed above the robotic arm. Participants were seated such that their elbow, pointing down to the ground, formed an angle of approximately 90°and their forehead was placed on a soft cushion attached to the frame of the setup. A virtual reality display was placed above the handle to allow participants to interact with virtual targets and obstacles. A white dot with a radius of 0.5 cm corresponding to the position of the participant's hand was displayed on the display throughout the whole experiment.

2.3 Experiment 1

In the Experiment 1 (see Figure II.1A), participants (N=14) were instructed to perform reaching movements to a visual target that was either a narrow square (2.5 cm x 2.5 cm) or a wide rectangle (30 cm x 2.5 cm) located 25 cm in the y-direction from the square home target (2.5 cm x 2.5 cm). The wide axis of the rectangle was aligned with the x-axis and was orthogonal to the straight-line path from the home target to

the target goal. Briefly, participants had to put a hand-aligned cursor on the home target (displayed as a red square). When the home target was reached, it turned green and after a random time delay (anywhere from 2-4s), the goal target was projected and participants could initiate their movement whenever they wanted. There were thus no constraints on the reaction time. Following exit from the home target, participants had to complete their movements between 350ms and 600ms from the start of movement in order to successfully complete the trial. The trial was successfully completed if participants reached the goal target within the prescribed time window and were able to stabilize the cursor in it for 500ms (from their arrival in the target). The goal target turned green at the end of a successful trial. For trials in which the timing window was not respected, the goal target turned blue or red to indicate that the movements were too fast or too slow, respectively.

Two types of perturbations could occur during movement. The first was a mechanical perturbation load consisting of a lateral step force applied by the robot to participant's hand. The magnitude of this force was $\pm$ 9N aligned with the x-axis, with a 10ms linear build-up. This force was triggered when the hand-aligned cursor crossed a virtual line parallel to the x-axis and located at 8cm from the center of the home target (see Figure II.1A). The step force was switched off at the end of the trial. The second type of perturbation that could occur was a visual perturbation whereby the shape of the target instantaneously changed from a narrow square to a wide rectangle, or vice versa. This switch in target shape was also triggered based on a position threshold, located either 2 cm (early target switch) or 8 cm (late target switch) from the center of the home target. Both perturbed and unperturbed trials were randomly interleaved such that participants could not predict the occurrence of either visual or mechanical perturbations.

Participants started with a training block of 20 trials in order to become familiar with the task and the force intensity of the perturbation loads. After completing this training block, six blocks of 72 trials were conducted. Each 72-trials block included: 24 trials involving no force or target switch, 16 trials with a mechanical perturbation only, and 32 trials with both visual and mechanical perturbations. Therefore, participants completed 432 trials, amongst which 24 trials which included each perturbation combination were included [e.g., direction of the mechanical perturbation changed (e.g., rightward versus leftward) and

timing of the target changed (e.g., early versus late)]. Participants were verbally encouraged to take advantage of the width of the target when it was a rectangle. To motivate participants, a score corresponding to their number of successful trials was projected next to the goal target.

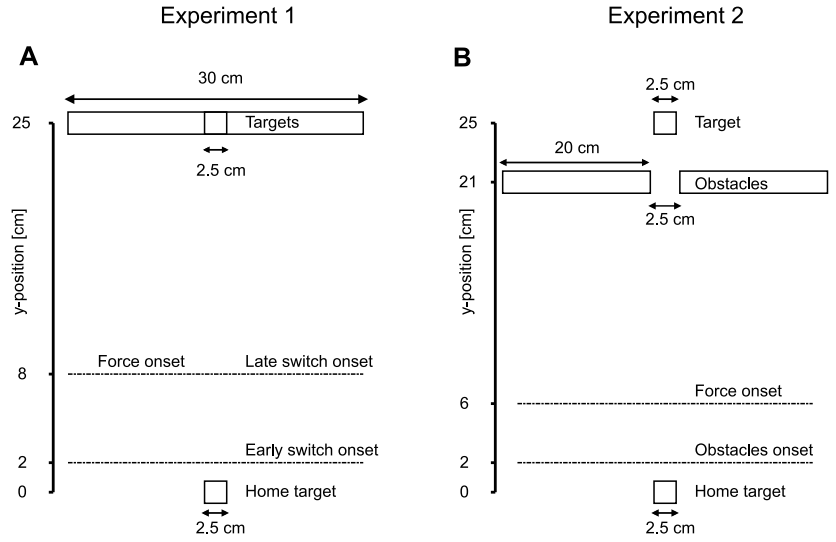

Figure II.1: Experimental paradigm. **A**. Schematic of Experiment 1, participants were asked to perform reaching movements from the home target to the goal target (square or rectangle). The dashed lines represent the y-position at which an early switch in target shape, a late switch in target shape, a force (perturbation load) were introduced. To succeed the trial, participants had to reach and stabilize in the goal target in the prescribed time window. **B**. Schematic of Experiment 2, participants were asked to perform reaching movements from the home target to the goal target while avoiding obstacles, if present. The dashed lines indicate the y-position at which obstacles or a mechanical force (perturbation load) were introduced. To succeed the trial, participants had to reach the goal target in the prescribed time window without hitting the obstacles.

2.4 Experiment 2

The second experiment was a variant of the first one and was designed to assess its reproducibility in a distinct context. Thus, instead of changing the structure of the goal, obstacles were introduced (see Figure II.1B). Briefly, participants (N=17) had to perform reaching movements towards a square target (2.5 cm x 2.5 cm) which was positioned similarly to Experiment 1. However, two virtual obstacles (20 cm x 2.5 cm each) were located 2.5 cm apart on either side of the target reach path, with their main axes aligned with the x-axis and orthogonal to

the main reaching direction. When participants hit these obstacles, a reaction force pushed their hand back to provide haptic feedback and they lost a point. Similar to Experiment 1, participants had to reach the home target to initiate the trial. After a random time delay, the goal target appeared as a red square and they could start their movement whenever they wanted. Again, two types of perturbations were applied during a movement. The first perturbation was the same mechanical perturbation used in Experiment 1, and it was triggered 6 cm from the center of the home target and was aligned with the x-direction (positive or negative). We observed in some pilot testing that, in this context, the success rate was very low if the mechanical perturbation was triggered at 8 cm from the center of the home target. For this reason, we decided to trigger this perturbation at 6 cm. The second type of perturbation consisted in the appearance and disappearance of obstacles. This perturbation was triggered when the hand of the participant crossed a virtual line located at 2 cm from the center of the home target and aligned with the x-axis. All trials were randomly interleaved. Following a similar procedure as Experiment 1, participants first performed a training block of 20 trials involving both force and obstacles changes. After completing this training block, participants performed six blocks of 72 trials. Each 72-trials block included: 32 unperturbed trials, 16 trials with a mechanical perturbation but no obstacles, and 24 trials which contained both mechanical (e.g., leftward or rightward) and visual perturbation with an appearance or disappearance of obstacles. Participants performed a total of 432 trials, including 24 trials which included each perturbation combination (e.g., direction of the mechanical perturbation changed (e.g.,rightward versus leftward) and appearance or disappearance of obstacles). Also, similar to Experiment 1, the number of successfully completed trials was projected next to the initial target to help motivate the participants.

2.5 Experiment 3

The third experiment was a control experiment designed to verify that the changes in behavior observed in Experiment 1 could be elicited without mechanical perturbations and that therefore it was not a consequence of the perturbation load applied to the arm specific to feedback control. Thus, this control experiment was a variant of Experiment 1, during which there was no mechanical perturbation. Similar to Ex-

periment 1, participants (N=8) had to reach for either a narrow square or a wide rectangular target located 20 cm in the y-direction from the home target. The timing constraints were the same as in Experiment 1: between 350 and 600ms to complete movement once the hand-aligned cursor exited the home target. During movement, changes in target structure as those of Experiment 1 could occur: the shape of the goal target could instantaneously change from a square to a rectangle or vice-versa. This change was triggered when participant's hand crossed a virtual line located at 5 cm from the home target. Movements with or without target switch were randomly interleaved and the visual perturbations occurred in one of three trials. Participants first performed a training block of 20 trials without target switch, followed by three 60-trials blocks consisting of 40 trials without online target switch, 10 trials with online switch from square to rectangle, and 10 trials with switches from rectangle to square per block.

In all experiments, mechanical an visual perturbations were triggered based on a position threshold in an attempt to reduce variability in joint configuration and electromyography data (EMG). Both kinematics and EMG data were aligned according to the onset of mechanical perturbation prior to data analysis. Since the early switch condition and the appearance/disappearance of obstacles were not triggered at the same position threshold as the mechanical perturbation, there was time jitter in the perturbations onset. Therefore, the early target switch was triggered 82.45 ± 5.11ms before the force onset. The late switch condition was triggered at the same time as the mechanical load. As a result, time onset was fixed with respect to the onset of the mechanical perturbation. However, since visual processing of the setup was necessary, there was a 40ms delay between triggering of the visual change (namely the change in target shape or the appearance/disappearance of obstacles) and the visual change on the virtual reality display due to hardware limitations. In this chapter, this 40ms delay was taken into account in all the analyses, which means that the onsets of visual changes (both change in target and in environmental context) reported have been delayed by 40ms with respect to their respective triggers.

2.6 Data recording

Raw kinematics data were sampled at 1 kHz and low-pass filtered with a 4^{th} order double-pass Butterworth filter (cut-off frequency, 20 Hz). Hand velocities along the x- and y-axes and accelerations along the x-axis were computed from numerical differentiations of the position data by using 4^{th} order centered finite differences. The raw EMG data were band-pass filtered with a 4^{th} order double-pass Butterworth filter with a band pass between 20 and 250 Hz. EMG data were normalized for each participant to the average activity collected when participants maintained postural control at the home target against a constant force of 9N. Data from the Pectoralis Major were normalized to the average activity in the same muscle while maintaining postural control against a rightward 9N force whereas data from Deltoid Posterior were normalized to the average activity recorded while maintaining postural control against a leftward 9N force. We measured the muscle activity from 500 to 2000ms after force onset in the calibration trials. Analysis of EMG responses was performed after aligning the trials to the force onset. Data processing and parameter extractions were performed by using Matlab 2016b software. Mixed model analyses were performed by using R (version 3.4.1) and the package nlme. The experimental condition was considered as a fixed effect while participants were considered as random offsets. These models were fitted by maximizing the log-likelihood and significance was considered when the p-value of the t-test on the fixed effect was smaller than 0.05 for the first significant time bin and 0.005 for the following time bins.

In the first two experiments, surface EMG electrodes (Bagnoli surface EMG Sensor, Delys Inc., Natick, MA, USA) were used to record muscle activity during movements. Based on previous studies ([Lowrey et al., 2017],[Nashed et al., 2012]), we concentrated on the Pectoralis Major (PM) and the Posterior Deltoid (PD). Indeed, recent studies using the same configuration reported strong recruitment of these muscles ([Crevecoeur et al., 2020a, Crevecoeur et al., 2014]). These muscles are stretched by the application of the lateral forces used in our experiments, and therefore, they are largely recruited for the feedback responses. The skin of each participant was cleaned and abraded with cotton wool and alcohol before electrodes were applied. Conduction gel was subsequently spread on the electrodes to improve signal quality. EMG data were sampled at a frequency of 1 kHz and

were amplified by a factor of 1000. In both experiments, a reference electrode was attached to the right ankle of the participant. No EMG data were collected during the third experiment.

During the first two experiments, we also collected EMG activity from other muscles for exploratory analyses. These additional muscles included the Triceps Lateralis and Brachioradialis during the Experiment 1, and the Anterior Deltoid, Teres Major, Infraspinatus, and Biceps during Experiment 2. For these other muscles, the patterns of response were too variable across the participants, likely because the joint configuration was not precisely controlled. Thus, the results described below concentrate on the two muscles for which a clear stretch response could be measured, as well as correlates of changes in behavior as measured with movement kinematics.

2.7 Statistical analysis

Final position of participant's hand were determined as the hand position when the total speed dropped below 3cm/s. Median values of end-point distributions were analyzed for statistically significant differences by using two samples Kolmogorov-Smirnov tests. Distributions obtained under early and late conditions were compared with distributions obtained without changes in target shape for the same force. Two-samples Kolmogorov-Smirnov tests were applied to compare the cumulative density functions of end-point distributions for the participants' hands at the end of the trial to identify statistically significant differences. To compare hand paths, the Peacock test was applied ([Peacock, 1983]). This multidimensional expansion of the two-samples Kolmogorov Smirnov test was used to determine whether two multidimensional distributions significantly differed. The null hypothesis of this test is that the two sets of multidimensional data points come from the same distribution. It is similar to the univariate distribution test, although it is performed in the two dimensions of the workspace and the largest difference is used to assess whether the samples differ. To perform this test, we implemented the algorithm developed by Xiao [Xiao, 2017].

To assess feedback responses for conditions with or without a change in target, normalized EMG data of the agonist muscles were used: for each perturbation direction, we only analyzed the stretch response in

the agonist muscle as the antagonist response in the shortened muscle did not show any difference. For each participant, their EMG data were averaged across trials and across time with ten non-overlaying bins of 25 ms width. The width of the bins has been determined based on prior testing in which we tested bins of 15 ms, 25 ms and 40 ms width. We observed qualitatively very similar results for the three tested bin sizes but noticed that small bins are more sensitive to noise whereas wide bins act as a low pass filter that could either over or underestimate onset times. Therefore, we selected a bin width of 25 ms that gave us a good trade-off between sensitivity and accuracy. Moreover, this width corresponds to the bin widths selected for similar analyses in previous studies [Kurtzer et al., 2008, Nashed et al., 2012]. The first and last 25 ms bins started 50 ms and 275 ms after the onset of force, respectively. Next, we applied a mixed linear model to the time average of the EMG data in each of these bins (see equation II.1). The trial condition, x_i, was the fixed effect factor (slope) and participants, z_j, were the random effect factor (offset). We assessed the difference between conditions by looking at the significance of the fixed effect parameter β_1. The random effect parameter was introduced to factor the inter-subject variability in EMG activation.

$$y_{ij} \approx \beta_0 + \beta_1 x_j + \gamma_j z_j + \epsilon_{ij} \tag{II.1}$$

Individual assessments of the effect of the condition for each participant were performed using Wilcoxon sign ranked tests. For all of the statistical tests performed, a significance level of p = 0.005 was used, as previously proposed [Benjamin and Berger, 2018]. To reduce the rate of false negative results, p-values between 0.05 and 0.005 were considered to be indicative of a significant trend. A threshold value of 0.05 was used to determine the onset of differences in the sliding analysis on the basis that bins exhibiting significant differences at this level were followed by bins exhibiting highly significant differences ($p < 0.005$). Effect size for binned EMG analyses were quantified using *Cohen's d* defined as the standardized mean difference between two groups of independent observations [Lakens, 2013].

3 Results

3.1 Experiment 1

To investigate whether a change in the shape of the goal target during a reaching movement would elicit a change in behavior, participants were instructed to reach to a target that was either a narrow square or a wide rectangle of which shape could switch during movement (see section 2.3). During these movements, a lateral step force could be applied to the hand of participants in order to enhance the effect of a change in target shape on control policy online. Changes in control policy were assessed based on movement kinematics and evoked EMG responses of the muscles stretched by the perturbation loads.

3.1.1 Kinematics

Consistent with previous results [Lowrey et al., 2017, Nashed et al., 2012], movement kinematics depended on the shape of the goal target. For instance, the end-point distribution of hand positions along the x-axis (Figure II.2A) was wider for the rectangle target (0.84 ± 2.05 cm) than for the square target (-0.11 ± 0.74cm) (Kolmogorov-Smirnov test, $p < 0.005$, $K = 0.3780$). This difference was observed for all of the participants, and individual differences were observed for 9 of the 14 participants. Traces of unperturbed trials for a representative participant are shown in Figure II.2A. For all of the participants, the reaching time for the small square target (median 422.3ms) was longer than that for the large rectangle target (median 389ms) (Wilcoxon signed-rank test, $z = 9.64$, rank sum = $9.89 * 10^5$, $p < 0.001$). Individual differences in reaching time were observed for 13 of the 14 participants.

Similarly, when perturbation loads were applied without a change in target shape, kinematics of the movement were found to depend on the shape of the target (Figure II.2B). Indeed, the final x-position of the participants' hands in trials involving rightward perturbation loads for the square target (1.04 ± 1.58 cm (median $\pm$ SEM)) was different from those to the rectangle target (7.41 ± 4.05 cm) as assessed by the Kolmogorov-Smirnov test ($p < 0.005$, $K = 0.5089$). This difference was observed for all participants (Figure II.2 C). Similar differences were

observed when leftward perturbations occurred (Figure II.2B). However, no differences in hand positions were observed along the y-axis (Figure II.2D).

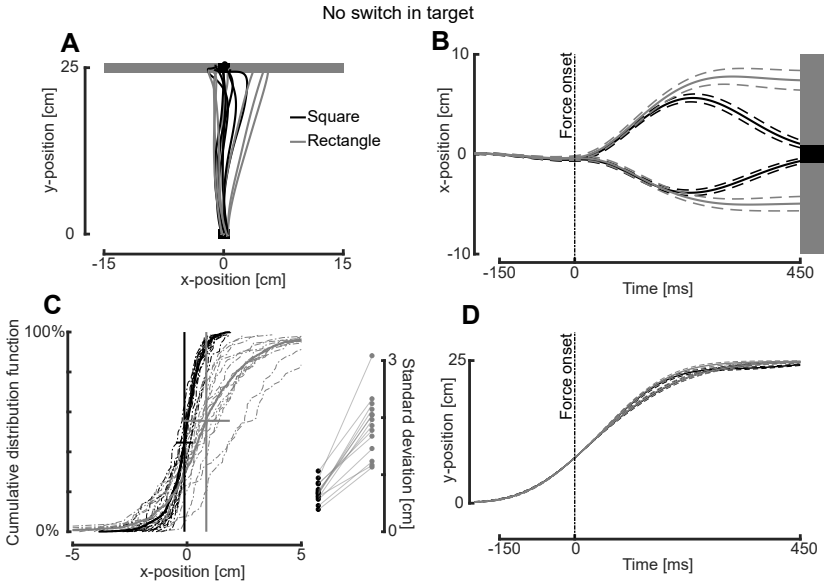

Figure II.2: Experiment 1 - Kinematics in the non-switching condition **A**. Hand paths of a representative participant toward a square (black) or rectangle (gray) target without any switch in target shape or perturbation load. **B**. Group mean (solid line) and SEM (dashed line) traces of the hand paths along the x-axis for participants reaching for a square (black) or a rectangle (gray) with perturbation loads applied. The onset of force is indicated with a vertical dashed line. **C**. Left: Cumulative distributions of hand end-points for the participants reaching to a square target (black) or a rectangular target (gray) without application of a perturbation load. Solid lines represent mean distribution values, while dotted lines represent individual distribution values. Right: Standard deviations of the individual end-point distributions of participants' hands reaching to a square (black dots) or rectangle (gray dots) target. **D**. Group mean and SEM y-position values of participants' hand path when reaching for a square (black) versus rectangle (gray) target with perturbation loads applied. The onset of force is indicated with a vertical dashed line.

These results confirmed previous findings showing that control policies exploit target redundancy to modulate feedback responses along dimensions that are not penalized by the target structure. Based on the observation that control policies used during movement depend on the shape of the goal target, we next wanted to examine whether similar adjustments would be evoked online when the goal target suddenly changed. Therefore, a change in the goal target shape from a square to a rectangle, or vice-versa, was introduced just after exiting the home

target (early switch) or just after the application of a mechanical perturbation load (late switch). End-point distributions as well as lateral acceleration profiles were investigated to assess the effect of this sudden change on the control policies.

When a target switch from square to rectangle occurred with rightward perturbation (Figure II.3A and B), the final hand positions along the x-axis for the early (6.01 ± 3.54 cm) and late (2.65 ± 2.68 cm) conditions were both different from those observed in the non-switching condition (1.04 ± 1.58 cm). These differences were significant for both early (Kolmogorov-Smirnov test, $p < 0.005$, $K = 0.3155$, red versus black in Figure II.3B) and late (Kolmogorov-Smirnov test, $p < 0.005$, $K = 0.6905$, blue versus black in Figure II.3B) conditions. The final position of the participants' hand also exhibited dependency on the timing of the target switch. Indeed, the comparison between early and late target switch (Figure II.3B, red and blue) revealed significant difference (Kolmogorov-Smirnov test, $p < 0.005$, $K = 0.4435$). Differences were also observed in the final hand positions along the y-axis. However, the effect size of these differences was five times smaller than the ones observed along the x-axis. Moreover, the stopping criterion based on 3 cm/s is met for a large proportion of trials in all conditions (more than 90% of the trials), which confirms a selective modulation of feedback gains.

Differences in final hand positions were also observed for the switches from rectangle to square targets. Both early and late switching conditions showed smaller final deviations of the participants' hand positions along the x-axis compared to the non-switching condition (see Figure II.3E for results obtained with rightward perturbation). Moreover, the early and late conditions showed different end-point distribution of participants' hands: the early switch condition showing smaller lateral deviation than the later one.

Similar observations characterized the responses to the target switches with leftward perturbations. Differences in the final x-position of the participants' hands between both switching conditions and the non-switching condition were also observed with leftward perturbation loads and with both target switches.

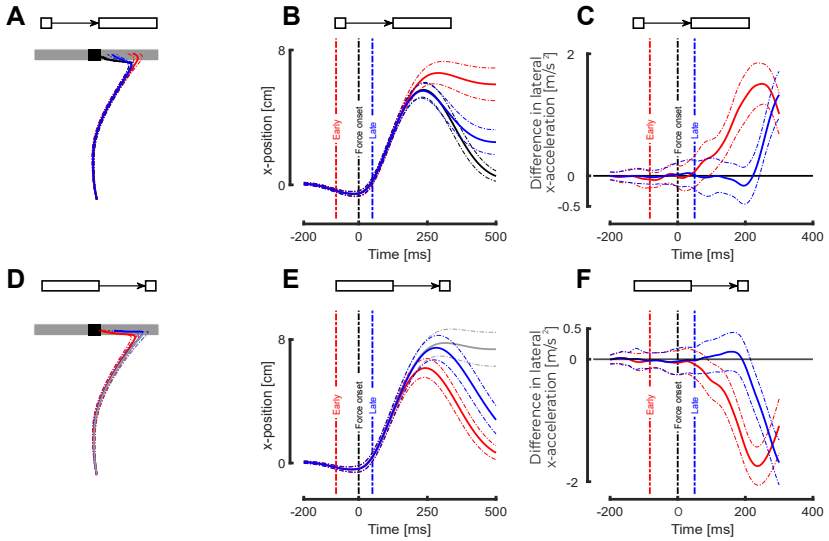

Figure II.3: Experiment 1 : Kinematics in the switching conditions **A**,**D** Mean (full lines) and SEM along the x-axis (dashed lines) of participants' hand paths as they reached toward a square (black) that could switch into a rectangle (**A**) early (red) or late (blue) and vice-versa (**D**) with rightward perturbation load. **B**,**E** X-position of participants' hands with respect to time as they reached toward a square (black) that turned early (red) or late (blue) into a rectangle (**B**) or vice-versa (**E**). The dashed lines represent the onset of the visual and mechanical changes. **C**,**F** Differences in lateral hand acceleration between the early or late conditions and the non-switching ones for a target that was initially a square (**C**) or a rectangle (**F**).

The lateral hand acceleration along the x-axis also exhibited clear dependence on the target switch. Figures II.3C and F represent the difference in lateral hand acceleration between the early (red) or late (blue) conditions and the corresponding non-switching conditions with rightward perturbations when the target turned from square to rectangle (Figure II.3C) or vice-versa (Figure II.3F). The increase in difference represented in Figure II.3C characterizes the decrease in corrective responses induced by the target switch that results in larger final hand deviations. Conversely, the decrease in difference of accelerations represented in Figure II.3F characterizes the increase in corrective responses that results in smaller lateral deviations. We determined the onset of these differences using a sliding Wilcoxon sign ranked test. The onset was defined as the first time where $p < 0.005$ at population level. When the target turned from square to rectangle with rightward perturbation

we found this onset at 61 and 236 ms after force onset for early and late switches, respectively. Similar onset times were observed for the leftward perturbations and for switches from rectangle to square.

Thus, online changes in target shape (from square to rectangle and vice-versa), elicited changes in movement kinematics for both early and late switch conditions. We observed that the kinematics reflected a relaxation in the corrective responses when the task constraints were decreased (square to rectangle) and an increase in corrective responses when the task constraints were increased (rectangle to square). Based on these observations, we next focused on the EMG data to gain insight into the latency of the changes in neural control evoked by switching of the target.

3.1.2 Muscle activity

Based on the behavioral analyses presented above, we hypothesized that evoked EMG responses to perturbation loads would increase or decrease when the change in goal target increased or decreased the task constraints, respectively. A key aspect was to assess when a change in EMG activity could be observed. Therefore, we initially examined global changes in agonist EMG responses consistent with the changes in end-point distributions based on the activity averaged across different time windows after introduction of a perturbation load. Then, we used a sliding window analysis to measure the latency at which a change in the goal altered the EMG responses to perturbation. For this purpose, we examined stretching induced by rightward and leftward perturbations in the Pectoralis Major and Posterior Deltoid, respectively.

We observed that the EMG responses to mechanical perturbations were tuned to the structure of the goal target. In Figures II.4A and B, the responses of PM and PD to rightward and leftward perturbations are respectively shown for the non-switching conditions. In both perturbation directions, we observed higher EMG responses in the stretched muscle when the participants were reaching for the square target goal rather than for the rectangular target goal (PM for rightward and PD for leftward perturbation). A mixed model analysis employing bins of 25ms was used to characterize these differences (see 2.7). For rightward perturbations, the first significant larger PM response associated with the square target was observed between 125 and 150ms ($t(657) = -2.1446$,

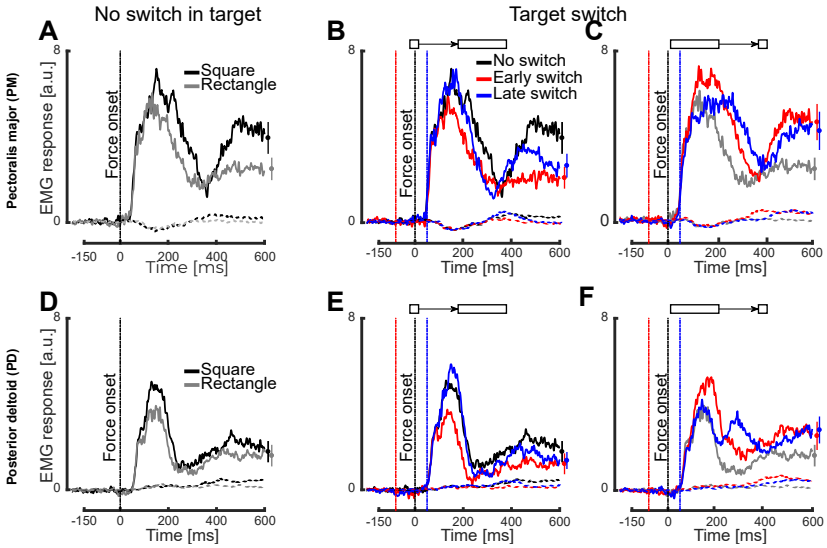

Figure II.4: Experiment 1 : EMG responses to perturbation load for the agonist (full lines) and antagonist (dashed lines) muscles **A**-**F**. Mean baseline reduced EMG responses in the PM (**A**-**C**) and PD (**D**-**F**) of participants with a rightward (**A**-**C**) or leftward (**D**-**F**) perturbation load (indicated as a vertical dashed line) applied while reaching to a square (black), a rectangle (gray), to a square switched to a rectangle (**B**,**E**) or vice versa (**C**,**F**), either early (red) or late (blue). For each plot, mean $\pm$ SEM values for the EMG responses 600ms after the onset of force are shown at the far right. In each panel, the mean baseline reduced EMG responses in the antagonist muscle (**A**-**C** PD and **D**-**F** PM) are shown in dashed lines.

$p = 0.0323$, $d = 0.19$) (Figure II.4A) followed by strongly significant differences in the subsequent time bins (all $p < 0.0005$). Thirteen of the 14 participants exhibited individual differences in trial distributions. For leftward perturbations, the first significant larger PD response associated with the square target was observed between 100 and $125ms$ ($t(657) = -4.0612$, $p < 0.005$, $d = 0.43$) (Figure II.4D), also followed by strongly significant differences (all $p < 0.005$). Eleven of the 14 participants exhibited individual differences in trial distributions. These differences occurred slightly later than those reported in previous work [Cross et al., 2019, Nashed et al., 2012, Lowrey et al., 2017], which might be attributable to the spatial layout, joint configuration and perturbation magnitude. A visual inspection indicated that this trend started in the long-latency epoch ($50 - 100$ms) without reaching significance.

Based on the observed EMG responses to perturbation loads for the square and rectangle targets separately, we hypothesized that a decrease in EMG response would be observed when the target switched from square to rectangle, whereas an increase in response would be observed for a switch from rectangle to square. When the target was switched from a square to a rectangle, we observed reduced EMG responses to perturbations for both PM and PD when they were stretched by the perturbation (Figures II.4B and E). When both early (red) and late (blue) conditions were compared to the non-switching condition (black), it appeared that the EMG response to perturbation was adjusted during movement. Indeed, when an early switch associated with rightward perturbation occurred, a decrease in PM response was observed for all of the participants between the $125 - 150$ms bin ($t(657) = -2.0591$, $p = 0.0399$, $d = 0.26$) and the $275 - 300$ms bin ($p < 0.0005$ for all of the intermediate bins). Five of the 14 participants exhibited individual differences in trial distributions. When a late switch and a rightward perturbation occurred, a decrease in the PM EMG response was observed across all of the participants between the $200 - 225$ms bin ($t(657) = -2.5964$, $p = 0.0096$, $d = 0.17$) and the $275 - 300$ms bin ($p < 0.005$ for all of the intermediate bins). Seven of the 14 participants exhibited individual differences in trial distributions. Meanwhile, for PD, an early switch condition associated with leftward perturbation resulted in a decreased EMG response between the $100 - 125$ms bin ($t(657) = -5.1381$, $p < 0.005$, $d = 0.3$) and the $250 - 275$ms bin ($p < 0.005$ for all of the intermediate bins). Seven of the 14 participants exhibited individual differences in trial distributions. However, under the late switch condition, no decrease in EMG response was observed for PD.

Conversely, when the target was switched from rectangle to square, PM and PD EMG responses to perturbation loads increased (Figures II.4C and F) when the muscles were stretched. The early (red) and late (blue) switch conditions were each compared to the non-switching condition (gray). For PM, the early switch condition associated with rightward perturbation resulted in an increase in EMG response from the $100 - 125$ms bin ($t(657) = -2.4266$, $p = 0.0155$, $d = 0.18$) to the $275 - 300$ms bin ($p < 0.0005$ for all of the intermediate bins). Nine of the 14 participants exhibited individual differences in trial distributions. When a late target switch was introduced, a significant increase in EMG response was observed from the $150 - 175$ms bin

($t(657) = -2.6505$, $p = 0.0082$, $d = 0.19$) to the $275 - 300$ms bin ($p < 0.005$ for all intermediate bins). Eleven of the 14 participants exhibited individual differences in trial distributions. For PD, the early switch condition associated with leftward perturbation elicited an increase in EMG response from the $100 - 125$ms bin ($t(657) = -3.2511$, $p = 0.0012$, $d = 0.19$) to the $275-300$ms bin ($p < 0.005$ for all intermediate bins). Ten of the 14 participants exhibited individual differences. Under the late switch condition, significant differences were observed from the $200 - 225$ms bin ($t(657) = -3.7121$, $p = 0.0002$, $d = 0.21$) to the $275 - 300$ms bin ($p < 0.005$ for all intermediate bins) and 10 of the 14 participants exhibited individual differences in trial distributions.

The kinematics and EMG responses to perturbation loads both indicate that the participants adjusted their behavior during movement in response to a switch in target shape. These adjustments were such that the corrective responses to perturbation increased when the target switched from rectangle to square and decreased when the target switched from square to rectangle. These results also suggest that the time onset of those responses correlates with the onset of the target switch. Indeed, we observed that differences in both kinematics and EMG responses occurred earlier in the trials involving an early target switch than in those involving a late target switch. To better characterize the delay between the trigger of a target switch and the first differences in EMG responses, we subtracted the EMG responses obtained under non-switching conditions from those obtained under switching conditions. Thus, the remaining signals represented the responses to the target switch itself (Figure II.5), and the onset of differences in the EMG responses to the target switch was found to depend on the condition. For example, the onset of differences in EMG response was associated with the first 25ms bin in which the average EMG activity significantly differed from zero. Those bins are represented as shaded rectangles in Figure II.5 for the different target switching conditions (e.g., early or late switching from square to rectangle or from rectangle to square) and for the two muscles as they were stretched (Figures II.5A and B for PM; Figures II.5C and D for PD). We also computed the individual mean values of the differences in EMG responses for the bins between 200ms and 375ms after a late onset of force (dots on the right side of each panel in II.5). When the square target was switched to the rectangle target, the mean values were negative ($p < 0.005$ in both muscles and for both early

and late switch conditions). When the rectangle target was switched to the square target, these mean values were positive ($p < 0.005$ in both muscles and for both early and late switch conditions).

Thus, the results of Experiment 1 indicate that participants adjusted their control policy during movements when the goal target was switched from a square to a rectangle, or vice-versa. These adjustments were such that the corrective responses are tuned to the new task goal. Significant differences were also observed in both the kinematics and EMG responses to the two switching conditions. Delays between the trigger of the target shape switch and the first differences in EMG response spanned from 150ms to 200ms. An analysis of the individual muscle sample data further indicated that changes occurred as early as $125 - 150$ms after the target switch. We believe that this timing for the first differences in EMG is a meaningful number but we acknowledge that it should be taken with caution as there is some uncertainty linked to the bin width.

3.2 Experiment 2

In order to demonstrate reproducibility and to test similar principles in a different context, we performed a second experiment which was designed to be a variant of Experiment 1. Briefly, we instructed participants to perform reaching movements to reach a square target in the presence of two virtual, rectangularly-shaped obstacles (see section 2.4). A lateral force was applied (as described in Experiment 1) to enhance possible changes in control which would be evoked by the appearance or disappearance of the obstacles. Changes in control policy were assessed based on movement kinematics and evoked EMG responses of the muscles stretched by the perturbation. An analysis of the participants' movement kinematics when the obstacle conditions remained unchanged revealed clear differences between the trials with and without obstacles. First, both the hand paths (Figures II.6A-F) and lateral hand deviations significantly differed when obstacle conditions were changed (appearing vs. disappearing, Peacock test). Second, the differences in lateral accelerations reported in Figures II.6C and F confirmed that the online changes in context (appearance and disappearance of obstacles) induced changed in responses to the perturbation. These kinematics results confirmed that the control policy of ongoing movements is adjusted online depending on whether obstacles are present.

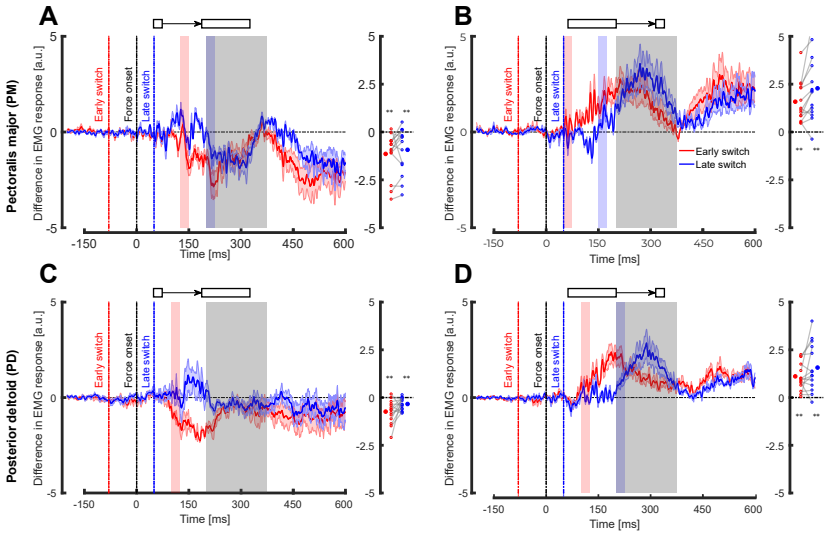

Figure II.5: Experiment 1 : Agonist EMG responses to visual change **A**-**D**. Mean and SEM differences in baseline reduced EMG responses in PM (**A**,**B**) or PD (**C**,**D**) while reaching to a square switched to a rectangle (**A**,**C**), or vice versa (**B**,**D**), either early (red) or late (blue). The onset of a rightward (**A**,**B**) or leftward (**C**,**D**) perturbation load is indicated with a black dashed line. The red and blue shaded rectangles in each panel represent the first significant bins for both conditions. At the far right of each panel, individual mean difference values in baseline reduced EMG are shown for the grey shaded rectangle for both early (red dots) and late (blue dots) conditions. The significance levels are $^{**}p < 0.005$ and $^{*}p < 0.05$ and these were assessed for all of the participants.

Similar to Experiment 1, we also examined evoked EMG responses of the PM (Figures II.7A-C) and PD (Figures II.7D-F) muscles as they were stretched by perturbations. Greater activity was detected in the presence (black) than in the absence of the obstacles (gray). Linear mixed model analysis revealed the first significant difference between conditions in the time bin $50-75$ms ($p < 0.005$) for the Pectoralis major. Figures II.7A and D present the responses for PM and PD, respectively. Differences were also observed in the evoked EMG responses to the perturbation loads when the obstacles appeared versus disappeared. For example, the first increase in EMG response for PM was observed when obstacles appeared during the $100 - 125$ms bin ($t(798) = 2.4816$, $p = 0.0133$, $d = 0.11$) and for PD during the $150 - 175$ms bin ($t(798) = 2.0009$, $p = 0.0497$,$d = 0.10$). Then, when the obstacles disappeared, the first decrease in EMG response in PM was observed during the $150 - 175$ms bin ($t(798) = -2.3674$, $p = 0.0181$, $d = 0.11$) and in PD during the

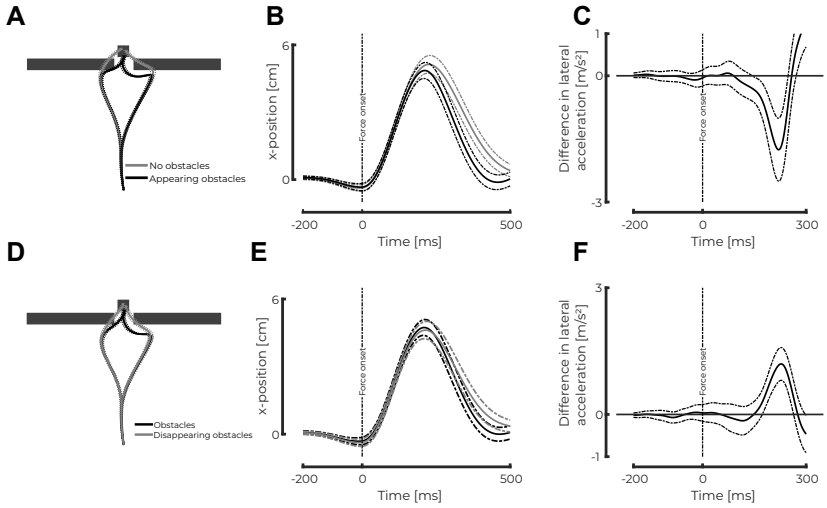

Figure II.6: Experiment 2 : Kinematics in the switching conditions **A**,**D** Mean (full lines) and SEM (dashed lines) of participants' hand paths as they reached toward a square without (**A**, gray) or with obstacles (**D**, black) that could appear (**A**, black) or disappear (**D**, gray). **B**,**E** x-position of participants' hands with respect to time as they reached toward a square with appearing (**B**, black) or disappearing (**E**, gray) obstacles with rightward perturbation loads. The dashed lines represent the onset of the visual and mechanical changes. **C**,**F** Differences in lateral hand acceleration between the switching conditions and the non-switching ones for appearing (**C**) or disappearing obstacles (**F**) with rightward perturbation loads.

$150-175$ms bin ($t(798) = 2.1805, p = 0.0295, d = 0.11$). These results in the presence and absence of obstacles are shown in Figure II.7B, E, C and F respectively. We further observed, similar to Experiment 1, that individual differences in trial distributions were detected for eight, seven, six, and seven participants, respectively.

Also similar to Experiment 1, data for the kinematics and stretched EMG responses to perturbation loads indicate that participants adjusted their policy during movement in response to changes in context (e.g., the appearance/disappearance of obstacles). To obtain an estimation of the delay between the appearance or disappearance of obstacles and the differences in EMG responses, we subtracted the EMG responses obtained when there were no changes in the obstacles from the EMG responses obtained when the conditions were changed. Specifically, for the trials involving the appearance of obstacles, the *no obstacles condition* was subtracted; for the trials where the obstacles disappeared, the *obstacles condition* was subtracted. The resulting signals corresponded

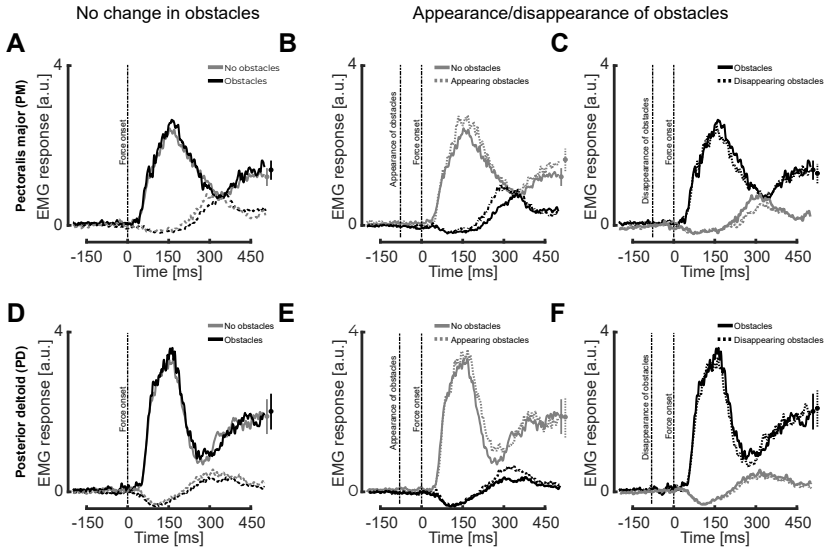

Figure II.7: Experiment 2 : EMG response to perturbation loads for agonist and antagonist muscles **A**-**F**. Group mean baseline reduced EMG response traces measured in the presence or absence of obstacles as indicated for PM (**A**-**C**) or PD (**D**-**F**) with a rightward (**A**-**C**) or leftward (**D**-**F**) perturbation load applied. Vertical dashed lines are used to indicate the time of force onset and the appearance or disappearance of obstacles as labelled. Group mean $\pm$ SEM values are shown at the far right of each plot. The corresponding antagonist response for each panel is shown as well.

to the impact of rapid changes in context on the stretched EMG responses (Figure II.8). Differences in EMG response occurred within the first 25ms in which a significantly non-zero average EMG activity was detected. The relevant bins are represented as shaded rectangles in Figure II.8. Observed delays between the onset of the appearance or disappearance of obstacles and the first differences in EMG responses spanned an interval of 180ms to 250ms.

3.3 Control experiment

In order to verify that the change in behavior observed in Experiment 1 could be elicited without mechanical perturbations (see section 2.4), we performed a third (control) experiment that was a variant of Experiment 1, in which there was no mechanical perturbation. Participants performed reaching movements to a goal target that could either be a small square or a large rectangle. During movement, the target shape could switch from square to rectangle and vice-versa. The occurrence

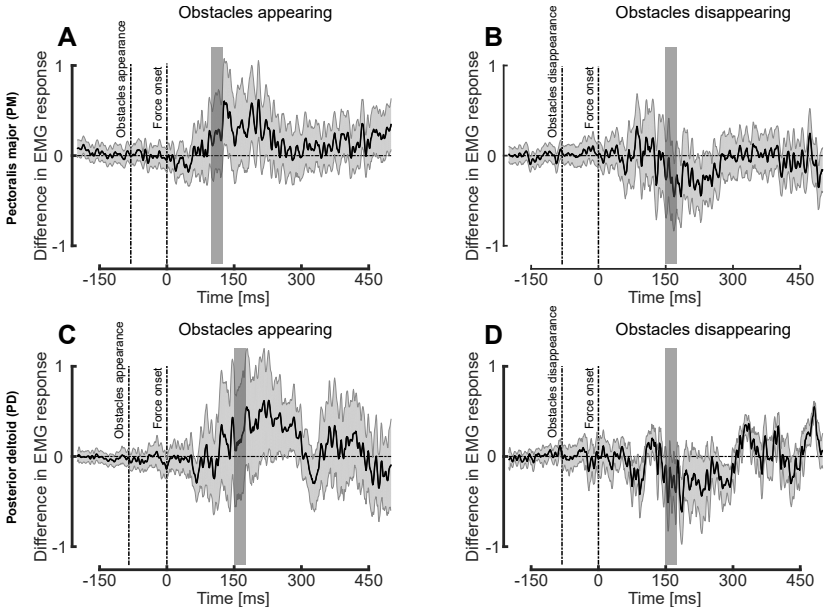

Figure II.8: Experiment 2 : Agonist EMG responses to visual changes **A**-**D**. Mean and SEM differences in baseline reduced EMG responses in PM (**A**,**B**) or PD (**C**,**D**) due to an appearance (**A**,**C**) or disappearance (**B**,**D**) of obstacles with rightward (**A**,**B**) or leftward (**C**,**D**) perturbation loads. The shaded rectangles in each panel represent the first time bin in which the response significantly differs from zero.

of a change in behavior induced by this switch in target shape was assessed based on hand kinematics. Figures II.9A and B represent the hand paths of a representative subject toward a square that turned into a rectangle and a rectangle that turned into a square respectively.

The end-point distribution of participants' hand for movements without target switch was different for square and rectangular targets as expected. The medians of these end-point distributions were not dependent on the target structure (Wilcoxon signed rank test, $z = 1.86$, ranksum $= 2.38 * 10^5$, $p = 0.0618$). However, we observed a wider distribution for rectangle (-0.7 ± 5.3cm) than for square targets (-0.03 ± 0.35 cm) across subjects (two-samples Kolmogorov-Smirnov test,$p < 0.005$, $K = 0.1125$). One-tailed paired t-test performed on individual variances confirmed this finding ($p < 0.005$, $t = -4.02$, $d = 1.18$). Strikingly, online target switches also elicited changes in end-point variance (Figure II.9C). Indeed, we observed an increase (one-tailed paired t-test: $p < 0.005$, $t = -4.10$, $d = -0.82$) or a decrease

(one-tailed paired t-test: $p < 0.005$, $t = 3.46$, $d = 1.06$) in end-point variance when the target switched from square to rectangle or from rectangle to square respectively Figure II.9D.

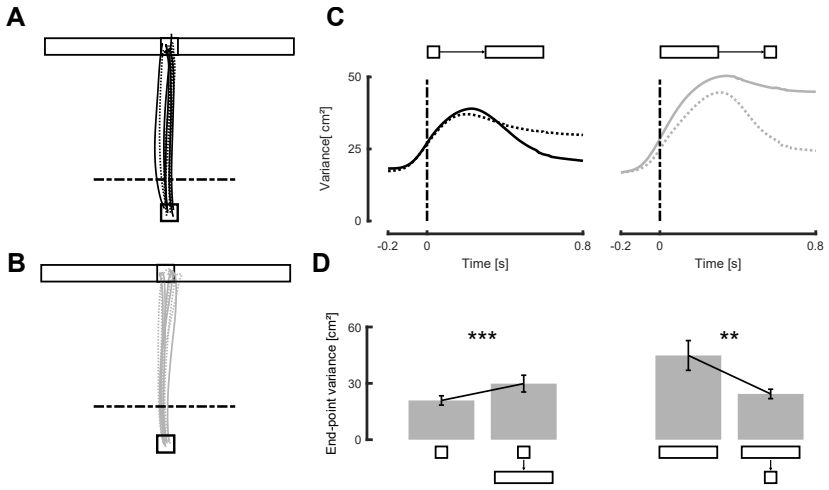

Figure II.9: Control experiment : Kinematics **A**-**B**. Hand paths of a representative participant toward a square (**A**) or rectangle (**B**) target, full lines correspond to the non switching condition and dashed lines correspond to the switching condition. Horizontal dashed lines represents the position of the onset of the visual perturbation. **C**. Evolution of the hand path variance along the x-axis for the unperturbed condition (full line) and the switching condition (dashed) for initial square (Left) and rectangle (Right) targets. **D**. Mean and SEM of the end-point variance of participants' hand position along the x-axis. The left bar plot represents the unperturbed condition and the right bar plot the perturbed position.

To conclude, online modification of the goal target structure or the environmental context elicited online changes in control, impacting corrections for lateral deviations due to natural variability (control Experiment) as well as feedback corrections to external loads applied during movements (Experiments 1 and 2).

4 Discussion

We conducted three experiments to determine whether and how quickly the control policy used by humans during reaching can be adjusted online when the structure of the target goal, or the environmental context, unexpectedly changed. In Experiment 1, we altered the structure of the goal target from a narrow square to a wide rectangle during movement

and vice-versa. In Experiment 2, obstacles located on both sides of the straight path to the target suddenly appeared or disappeared during movement. In both experiments, we found that the control policies used by participants adjusted online following the changes in target shape or context. In Experiment 1, the evoked EMG responses to mechanical perturbations adjusted in as little as 150ms following the change in target structure. Similarly, in Experiment 2, adjustments of the responses to mechanical perturbations were observed 180ms after the appearance or disappearance of obstacles. Finally, in Experiment 3, we observed a change in movement execution in response to the same target switch as in Experiment 1, even in the absence of mechanical perturbation. This shows that the evoked modulation of feedback gains is not only expressed or evoked by the occurrence of an important external disturbance, as induced by the perturbation loads, but instead it also reflects a general adjustment of control expressed in both perturbed and unperturbed movements. Specifically, this modulation of feedback gains reflects a selective correction of the perturbation loads explained by the change in target redundancy [Liu and Todorov, 2007].

This interpretation is in line with the theory of optimal feedback control (OFC) which suggests that the control policy used to perform movement and to correct for errors is tuned to the task goal and to its environmental context [Liu and Todorov, 2007, Scott, 2004, Todorov and Jordan, 2002]. Mathematically, the tuning of the control policy reflects the minimization of a cost-function, which captures the intended behavioral performance by weighting motor cost and penalties on motor errors (e.g. position, velocity, and force). In order to model the structure of the goal target in our experiments, a selective penalty on the x-coordinate for the square versus rectangle target could have easily been introduced. However, this modeling approach would be more complicated for the obstacle conditions since non-quadratic penalties would be involved but clearly context-dependent changes in control policy are expected.

In this framework, our experiments indirectly probed the two terms typically used in standard OFC models: a term related to behavioral performance that penalizes error in the state and a term related to motor cost that penalizes control commands. On the one hand, a switch from a rectangle to a square, or an appearance of obstacles, clearly alters the penalty related to the state. On the other hand, when the

target switched from a square to a rectangle, or when the obstacles disappeared, adjustments of control commands were not mandatory. The visual perturbation only modified the penalty on behavioral performance and left the penalty on motor cost unaffected, but the relative weight of motor cost likely promoted an adjustment of the control policy when the task became easier, as following a switch from a square to a rectangle, or when the obstacles disappeared. The net result is a decrease in EMG-evoked responses. This adjustment is observed with the rightward perturbation for both early and late target switches (Figure II.4B, red and blue traces, respectively), yet only for the early target switch with leftward perturbation (Figure II.4E, red trace). Moreover, even though the penalty regarding behavioral performance decreased, participants did not have to adjust their control policy to succeed in the trial. The fact that those adjustments are not mandatory, may explain why the decrease observed was not strong in all cases at short latencies, although an overall decrease later in the trial was always observed (right part of Figures II.5A and C).

Importantly, our study differs from previous research showing that target structure and size affected the control policy across conditions [Knill et al., 2011, Lowrey et al., 2017, Nashed et al., 2012, Scholz et al., 2000, Soechting, 1984]. These studies showed that target redundancy was handled during both motor planning and execution [Berret and Jean, 2016, Togo et al., 2017], and that feedback responses to mechanical perturbations were tuned to target redundancy [Lowrey et al., 2017, Nashed et al., 2012]. Our contribution was to show that the process of selecting a control policy adjusted to the target redundancy could also modify an ongoing movement. Our results could not be directly inferred from previous work on target [Dimitriou et al., 2013, Georgopoulos et al., 1981, Goodale et al., 1986, Gritsenko et al., 2009, Gritsenko and Kalaska, 2010, Mutha et al., 2008, Pélisson et al., 1986, Izawa and Shadmehr, 2008, Smeets and Brenner, 2003] and cursor jumps[Franklin and Wolpert, 2008, Sarlegna et al., 2003] because these perturbations did not necessarily implied a re-planning of movement. Indeed, it was recently demonstrated that both target and cursor jumps can be handled through state-feedback control using an infinite horizon optimal feedback controller [Li et al., 2018, Qian et al., 2013]. Thus, our key contribution was to probe changes in target structure and in environmental context to measure the latencies of online changes in control policies.

Besides the occurrence of online adjustments of the control policy, we also sought to measure the latency at which these adjustments occurred. In the first two experiments, the average latencies observed in the EMG responses were surprisingly short: 150ms in Experiment 1 and 180ms in Experiment 2. In comparison, visuomotor corrections by muscles in a target jump paradigms emerge within 100ms after the target jump [Day and Lyon, 2000, Franklin and Wolpert, 2008, Knill et al., 2011]. Thus, we conclude that it takes on average 150ms to re-plan the control policy of the ongoing movement, as revealed by changes in stretch responses to mechanical perturbations that occurred later than the target switch. This delay was longer than typical latencies of visuomotor reflexes [Franklin and Wolpert, 2008, Izawa and Shadmehr, 2008], and likely corresponded to early voluntary corrections. Interestingly, the difference of 30ms observed across Exp. 1 and 2 may be due to differences in the tasks performed. While the first change only affects the task goal, the second change affects the environmental context. A recent study also reported an additional 30ms latency in feedback responses to cursor jumps when the response was influenced by the presence of obstacles in the context of motor responses to visual perturbations [Cross et al., 2019]. Indeed, Cross and colleagues have shown that the motor response to a visual perturbation includes two distinct phases. The first phase starts after 90ms and is sensitive to goal redundancy. Meanwhile, the second phase starts after 120ms and is sensitive to environmental factors. Similarly, in the present study, a change in goal and a change in context resulted in a 30ms difference in latency. Thus, it is conceivable that 30ms is a common additional processing time for tasks involving environmental obstacles. Considering a fixed visuomotor delay of 100ms [Franklin and Wolpert, 2008, Saunders and Knill, 2004, Smeets and Brenner, 2003], and 30ms across experiments linked to a specific processing of obstacles, the measured latencies of target-dependent responses suggest that a change in control policy requires 50ms.

In order to set constraints on the online processes that adjust control policy, we compared the latencies observed in the present study with the visuomotor delays previously measured with target or cursor jumps and also with reported reaction times required to initiate a movement. The latencies we observed were longer than visuomotor delays associated with target or cursor jumps (100ms), which suggests that a more demanding process than the one involved in target

jumps is used. On the other hand, the latencies we measured were shorter than the reaction times characterized for initiating a movement [Georgopoulos et al., 1981]. The latter result suggests that the control policy is not re-computed from scratch. However, this assessment has to be taken with cautious since Haith and colleagues [Haith et al., 2016] showed that these reaction times not only reflect control policy selection and that movement preparation can occur in as little as 130ms. It is also important to stress out that the reaction times we used for these comparisons were mostly measured for transitions from postural control to reaching movements. In a previous study [Orban de Xivry et al., 2017], reaction times were reduced if movement planning and movement execution overlapped. This result is consistent with the present experiments since our participants had already started their movement when they had to adjust their control policy.

Two different kinds of adjustments in control policy could occur during movements in our experiments. The first is a re-computation of the control policy for the new task, corresponding to the theoretical framework of model predictive control [Lee, 2011]. Many recent theories of motor control consider that the control policy is computed before movement is initiated [Harris and Wolpert, 1998, Scott, 2004, Todorov and Jordan, 2002, Wong et al., 2015]. It has recently been suggested that specification of this control policy may overlap with movement execution [Orban de Xivry et al., 2017], which could support hypothesis about the process that perform the adjustment in control policy. The second process which could produce an online adjustment in control policy is a switch between different control policies stored in the brain. It has already been suggested that the central nervous system is able to change online the target goal of a movement if a perturbation interferes sufficiently with the task [Nashed et al., 2014]. Other recent studies have suggested that multiple motor plans may be handled in parallel by the central nervous system [Gallivan et al., 2016b, Gallivan et al., 2016a]. The present data do not allow us to determine whether the observed adjustments corresponded to a switching or re-computation. Thus, further studies are needed to elucidate the details of this process.

Interestingly, when the goal target was switched from a rectangle to a square in Experiment 1, we observed a shorter latency of the EMG-evoked motor response for the late switching condition than for

the early one. This difference may reflect the participants' urgency in correcting their movement toward the new goal target. Indeed, long-latency responses and early voluntary epochs of responses to perturbations have been shown to scale with task-related urgency [Crevecoeur et al., 2013]. In the present study, the participants' urgency to correct mechanical perturbations was larger when the switch from a rectangle target to a square target was made late in the trial. This result could explain the smaller latency of the response to the perturbation load. Another factor which may have potentially influenced these differences in latencies is the temporal evolution of feedback gains during reaching. It has been shown both experimentally and theoretically that feedback gains follow a skewed bell shape curve over time [Dimitriou et al., 2013, Liu and Todorov, 2007]. As a result, different responses to the same perturbation are applied at distinct points in time. In our experiment, the same differences were observed when the target was switched from a square to a rectangle.

Understanding the neural basis of the change in movement control remains an open question, however our developments provide a measurement of the latency and sets constraints on candidate underlying circuits. On the one hand, both spinal and supraspinal circuits are recruited during movement and following disturbances, but when the target did not change, the task-dependency emerged in the long-latency and voluntary epochs, consistent with previous work [Lowrey et al., 2017] and suggesting a contribution of cortical origin [Matthews, 1991, Pruszynski et al., 2011, Scott, 2004]. On the other hand, the main timing constraint in the condition where the target switched was due to the processing of the visual system. Rapid reflexive responses to visual perturbations are typically evoked in a little less than 100ms [Day and Lyon, 2000, Franklin and Wolpert, 2008, Knill et al., 2011], but we observed longer latencies for task-related changes likely due to more demanding processing associated with a change in control policy. In fact, our measure of the latency associated with changes in the shape of the target involves both visual and somatosensory systems, thus it is reasonable to speculate that associative regions in parietal cortex play a central role. With regards to timing, this suggestion is consistent with a cortical pathway collecting sensory signals from V1 and S1 to modulate the generation of motor commands through M1 [Omrani et al., 2016].

To conclude, we have demonstrated that humans are able to quickly update their control policy when a task goal or its environment changes during reaching movements, and that these changes characterize unperturbed movements as well as feedback responses to mechanical loads applied to their limb. We observed adjustment delays which were slightly longer than the visuomotor delays observed in a target jump paradigm, which suggests that a more demanding neural operation is needed to the change in control policy. Our estimate of the latency associated with this process is 30 ms, therefore allowing fast motor adjustments suitable for the execution of ongoing movements in a dynamic context. These valuable insights provide direction for further studies of movement planning and how it interacts with movement execution dynamically.

Chapter III

Dynamical changes in control policy during ongoing movements

How can we prepare for the future if we won't acknowledge the past ?

Nancy K. Jemisin

Humans are able to execute reaching movements in environments where unexpected changes of the limb positions or the task could happen. The control policies they use to execute these movements are tuned to the task goal and to the environmental context such that they can appropriately handle perturbations. These control policies can be adjusted online if an abrupt change in task demand occur. However, it is still unknown whether they can be dynamically adjusted within movements in presence of dynamical changes in task demands. Here we address this question by gradually changing the task demands demands during movement and investigating participants' behavior. More specifically, we studied whether the feedback responses to mechanical perturbations integrate dynamical changes in the structure of the goal target. We designed a task wherein the width of the goal target gradually decreased during movement to dynamically alter target redundancy. We assessed whether these dynamical changes in target structure altered the feedback responses to mechanical perturbations

based on kinematics data and surface EMG recordings. We observed that the feedback responses to the mechanical perturbations exhibited target-dependent modulations that reflected the dynamical changes in target width. Moreover, we reported that these modulations of the feedback responses depended on the rate of change in target width. In all, our results demonstrate that the control policies used by humans to perform reaching movements can be dynamically adjusted to the changes in task constraints. This suggests the existence of a continuous feedback mechanism that conveys task-related information to mechanism selecting the control policy during movement.

1 Introduction

From an early age, humans and animals have learned to evolve and execute movements in complex environments. Unexpected abrupt and gradual changes in these environments can interfere with their ability to execute and achieve movements. Humans have demonstrated the capability to rapidly adapt and respond to these changes. Indeed, previous studies reported that human participants can learn to interact with unknown mechanical [Krakauer et al., 1999, Lackner and Dizio, 1994, Shadmehr and Mussa-Ivaldi, 1994, Singh and Scott, 2003] or visual [Krakauer et al., 2000, Mazzoni and Krakaueur, 2006] perturbations, highlighting rapid and complex learning mechanisms in humans. They reported that participants were able to adjust their behavior, already within a few trials, to correctly respond to these perturbations even if they could not easily be modeled. Similarly, a large body of work investigated whether and how humans can respond to unexpected mechanical [Knill et al., 2011, Nashed et al., 2012, Lowrey et al., 2017] and visual [Georgopoulos et al., 1981, Soechting and Lacquaniti, 1983, Sarlegna and Mutha, 2014] perturbations. The goal of these studies was to investigate the feedback mechanisms underlying the control of reaching movements. Altogether, these studies demonstrate that humans have the capacity to integrate unexpected and unmodeled perturbations during movement as they evolve in complex dynamical environments.

Upper limb reaching movements can be modeled by the theoretical framework of Optimal Feedback Control (OFC). This framework posits that the control policies underlying reaching movements are selected such that they optimize a given performance index

[Todorov and Jordan, 2002]. This performance index is captured by a cost-function, consisting of a weighted sum of the motor cost and the motor performance (end-point error, velocity, etc). The optimal control policy, obtained through minimization of this cost-function, is therefore specifically tuned to the task goal as it can be assessed from behavior and EMG recording during unperturbed and perturbed trials [Nashed et al., 2012, Nashed et al., 2014, Lowrey et al., 2017]. For instance, goal-dependent modulation of feedback responses has been reported in studies that introduced visual and mechanical perturbations [Todorov, 2004, Scott, 2004]. This framework has been used to motivate many previously conducted studies that demonstrated that it could predict the behavioral outcomes of many perturbation paradigms [Diedrichsen, 2007, Izawa and Shadmehr, 2008, Omrani et al., 2013]. A caveat of this framework as it was initially formulated is that the cost-function and the corresponding optimal control policy are selected before the beginning of the movement. Therefore, if the task constraints are modified during an ongoing movement in such a way that the cost function should be updated, the selected control policy will not be tuned to the new task goal anymore.

A recent study demonstrated that the control policy used by humans to perform reaching movements can be adjusted online in response to changes in task constraints or environmental context [De Comite et al., 2021]. For this purpose, the authors abruptly altered the structure of the goal target during movement, from a narrow to a wide rectangle or vice versa. They reported rapid goal-dependent adjustments of the control policies within movement in 150ms following the change in target structure. These adjustments of control policies were coherent with the changes in task demands. Similar online adjustments could also be elicited by the sudden appearance or disappearance of obstacles. This study show that the control policies underlying reaching movements could be changed during movement, contrary to what the OFC framework predicts. However, it is still unknown whether similar online adjustments could occur if the change in task constraints evolves gradually, which is often the case in the real world.

Here we address this question in a behavioral experiment where participants had to perform reaching movements to a goal target in a dynamically changing environment. More specifically, we investigated whether and how participants adjust their control policy when the

structure of the goal target continuously changed within movements. The width of the goal target was gradually decreased as they were reaching to it, corresponding to a dynamical modification of the target redundancy along its main axis. In this context, the first alternative is that the control policy does not integrate the dynamical changes in target and remains constant throughout movement, leading to no changes in behavior and feedback responses to mechanical perturbations. The second alternative is that the dynamical changes in task constraints will evoke adjustments in the control policy reflected by modifications of the behavior and feedback responses that integrate the changes in target structure. The second alternative was the correct one as we observed that participants adapted their control policy online in these dynamical contexts. Strikingly, these dynamical adjustments were also modulated by the rate of change in target structure. We indeed observe that faster decreases in target width elicited larger feedback corrections. These results consist in an evidence for a continuous feedback mechanism that allows to dynamically adjust human motor control policies within movements.

2 Methods

2.1 Participants

Fourteen participants (10 females) ranging in age from 18 to 30 years old participated in this study. Participants were naïve to the purpose of the study, had normal to corrected vision, and had no known neurological disorder. The ethics committee of Université catholique de Louvain (UCLouvain) approved the experimental procedures and participants provided their written informed consent.

2.2 Experimental paradigm

Participants were seated on an adjustable chair in front of a Kinarm end-point robotic device (Kinarm, Kingston, ON, Canada) and grasped the handle of the right robotic arm with their right hand. The robotic arm allowed movements in the horizontal plane and direct vision of both their hand and the robotic arm was blocked. Participants sat such that, at rest, their elbow formed an angle of approximately 90° pointed downward and their forehead rested on a soft cushion attached

to the frame of the robot. A semi-transparent mirror, located above the handle and reflecting a virtual reality display (VPIXX, 120 Hz) allowed participants to interact with visual targets. A white dot of 0.5 cm radius aligned to the position of the right handle was presented on this display during the whole experiment.

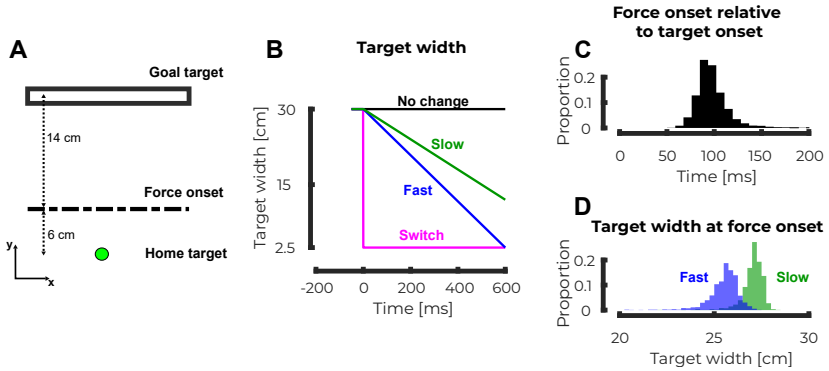

Figure III.1: Experimental paradigm: **A**. Schematic representation of the task paradigm. Participants had to perform reaching movements from the home target to the goal target, initially represented as a 30cm wide rectangle. During movement, they could experience mechanical step forces triggered in position at 6cm from the home target (dashed black line) and visual changes in target size (triggered as they exited the home target). **B**. Evolution of the target width with respect to time in the different target conditions. The time axis is aligned on the visual perturbation onset. **C**. Histogram of the distribution of the time span between the visual and mechanical perturbation onsets (resp. target and forces onset) across all participants and conditions. **D**. Histogram of the distribution of the target width at force onset for the two dynamical conditions (fast in blue and slow in green) across all participants.

In this experiment, participants (N=14) were instructed to perform reaching movements to a visual target initially represented as a wide rectangle (30x2.5cm) located at 20cm in the y-direction from the home target represented as a circle of 1.5cm of diameter. The main axis of the rectangle was aligned with the x-axis and was orthogonal to the straight-line path from the home target to the center of the goal target (see Figure III.1A). Participants first had to bring the hand-aligned cursor in the home target displayed as a red circle that turned green as they reached it. After a random time delay (anywhere from 1 to 2s), the goal target was projected as a gray rectangle and participants could begin their movement whenever they wanted. There were therefore no constraints on the reaction time. Following exit from the home target, participants had to complete their movement between 350 and 600ms in order to successfully complete the trial. The trial was successfully

completed if they reached the goal target within the prescribed time window and were able to stabilize the cursor in it for 500ms. The goal target turned green at the end of successfull trials and turned red otherwise.

During movements, two types of disturbances could occur. The first one was a mechanical perturbation load consisting of a lateral step force applied by the robot to participant's hand. The magnitude of this force was $\pm$9N aligned with the x-axis, with a 10ms linear build-up. This force was triggered when the hand-aligned cursor crossed a virtual line parallel to the x-axis and located at 6cm from the center of the home target (see Figure III.1A, horizontal dashed line). The step force was switched off at the end of the trial. The second type of perturbation that could occur was a visual perturbation whereby the target width could change as participants exited the home target (called *target condition* through this chapter). This change could either be an abrupt change (*switch* condition) to a small square of 2.5x2.5cm (Figure III.1B, magenta) or a gradual change in target width either at a speed of -30cm/s (*slow* condition, green in Figure III.1B) or at a speed of -45.8cm/s (*fast* condition, blue in Figure III.1B). The speed of the *fast* condition was selected such that the target width at the end of the movement[1] was the same as in the *switch* condition. The decrease in target width stopped as participants entered the goal target. Unperturbed and perturbed trials were randomly interleaved such that participants could not predict the occurrence of either visual or mechanical perturbations. Whatever the target condition was, participants were instructed to reach the target as it was presented: narrower target reduced the potential correct movements. Participants started with a 25-trials training block in order to become familiar with the task, the timing constraints, and the force intensity of perturbation loads. Importantly, this training block did not include any visual perturbation. After completing this training block, participants performed 6 blocks of 82 trials. Each 82-trials block contained: 38 trials without mechanical perturbation (20 with no target change and 6 for each target condition) and 44 trials with mechanical perturbation (20 with no target change and 8 for each target condition, equally probable for rightward and leftward mechanical perturbations). Participants performed a total of 492 trials, including 24 of each com-

[1]Since participants had a relatively wide time window to execute movement, the upper bound of this time windows has been considered for the correspondence.

bination of perturbations (direction of the mechanical perturbation and target conditions). A total score corresponding to the number of successful trials was projected next to the home target. Participants were compensated 15€ for their participation.

Since both visual and mechanical perturbations were triggered based on position threshold (respectively when participants exited the home target and when they crossed a virtual line located at 6cm from the center of the home target), there was some time jitter between the two perturbations. The distribution of this time jitter is represented in Figure III.1C and had a median value of 96±6.41ms. As the target width in the *slow* and *fast* conditions were continuously changing with time, some jitter was also present in the target width at mechanical perturbation onset. In the *fast* condition we observed a median value of 25.6±0.3cm while in the *slow* condition we observed a median value of 27.1±0.15cm represented in Figure III.1D respectively in blue and green.

2.3 Data collection and analysis

Raw kinematics data were sampled at 1kHz and low-pass filtered using a 4^{th} order double-pass Butterworth filter with cut-off frequency of 20Hz. Hand velocity, acceleration and jerk were computed from numerical differentiation of the position using 4^{th} order centered finite differences. Surface EMG electrodes (Bagnoli surface EMG sensor, Delsys INC. Natick, MA, USA) were used to record muscles activity during movements. We measured the Pectoralis Major (PM) and the Posterior Deltoid (PD) based on previous studies [Lowrey et al., 2017, De Comite et al., 2021] that showed that they are stretched, in this configuration, by the application of forces opposite to their actions, and therefore largely recruited by feedback responses. Before applying the electrodes, the skin of participants was cleaned and abraded with cotton wool and alcohol. Conduction gel was applied on the electrodes to improve the quality of signals. The EMG data were sampled at a frequency of 1kHz and amplified by a factor of 10 000. A reference electrode was attached to the right ankle of participants. Raw EMG data from PM and PD were band-pass filtered using a 4^{th} order double-pass Butterworth filter (cut-offs: 20 and 250Hz), rectified, aligned to force onset and averaged across trials or time windows as specified in section 3.

EMG data were normalized for each participant to the average activity collected as they maintained postural control against a constant force of 9N. Data from PM were normalized by the EMG activity in the same muscle while performing postural control against a rightward force whereas data from PD were normalized by the EMG activity in the same muscle while performing postural control against a leftward force. This calibration procedure was applied after the second and the fourth blocks.

2.4 Statistical analyses

Data processing and parameter extractions were performed using Matlab 2019a. We fitted linear mixed models to infer the effect of target conditions on different kinematics parameters and on the EMG activities. These models were fitted using the *fitlme* function of Matlab and the formula used was the following:

$$\text{Parameter} = \beta_0 + \beta_1 * \text{Condition} + \alpha_i \tag{III.1}$$

In this formula, the fixed predictors were the intercept (β_0) and the target condition (β_1) while participants were included as a random offset (α_i). For all linear mixed model analyses that we performed, we reported the estimate of β_1, the t-statistics for this estimate and the corresponding p-value. One-tailed Wilcoxon ranksum tests were used for post-hoc analyses where we collapsed data across trials and participants to compare the different conditions.

In order to investigate the time lag between the first and last trials in the dynamical conditions, we used a cross-correlation analysis applied on resampled data. We generated 1000 bootstrap samples from the individual acceleration profiles. For each of these samples, we computed the mean acceleration traces for the first and last trials and computed the cross-correlation between these two mean traces. We then extracted the peak value of this cross-correlation, corresponding to the lag between the two signals. The bootstrap resampling allowed us to obtain a distribution for this lag such that we could perform statistical analyses on it. Wilcoxon signed rank test was used to assess whether this time lag was significantly different from zero. In all our analyses, significance

was considered at the level of p=0.05 even though we decided to exactly report any p-value that was larger than p=0.005 as previously proposed [Benjamin and Berger, 2018].

3 Results

3.1 Gradual changes in target width induce dynamical adjustments in control policy

To determine whether gradual changes in target width impact the control policies used to execute movements, participants were asked to perform reaching movements to a target that was initially a 30cm wide rectangle, the width of which could change online. The target could either abruptly turn into a 2.5cm wide square target (*switch* condition) or gradually decrease in width either with a high (*fast* condition) or a low speed (*slow* condition). Unexpected mechanical perturbations were used during movements to elicit feedback corrections. Adjustments in control policies were assessed by investigating movement kinematics and EMG correlates of the shoulder flexor and extensor muscles.

3.1.1 Kinematics

We observed that the target condition influenced participants' behavior. Indeed, the mean hand path trajectories in the mechanically perturbed conditions (Figure III.2A) differed for the different target conditions. Consistently with our previous findings [De Comite et al., 2021] we observed online adjustments in the behavior in the *switch* condition (magenta) compared to the *no change* condition (black). These adjustments consisted of smaller lateral hand deviation in the *switch* condition (Figure III.2B and C, black and magenta traces). Interestingly, the behaviors in the dynamical target conditions (*slow* and *fast*, green and blue respectively) differed from both the *no change* and *switch* conditions. In order to quantify these differences, we investigated the maximal hand deviation induced by the mechanical perturbations as well as the final hand position defined as the x-position of the hand as its velocity dropped below 2cm/s.

The maximal lateral hand deviation induced by rightward mechanical perturbations (Figure III.2D) varied across target conditions. Indeed, a linear mixed model analysis (see section 2) revealed a significant effect of target condition (β_1=0.0417$\pm$0.0022, t=18.14, p<0.005) on the maximal hand deviation with larger deviations for slower changes in target width. Importantly, post-hoc analyses revealed that the maximal hand deviation was larger in the *no change* condition than in both dynamical conditions (*slow* one-tailed ranksum test z=10.36, p<0.005 and *fast* z=12.69, p<0.005). Moreover, this hand deviation was larger in these dynamical conditions than it was in the *switch* condition (*fast* one-tailed ranksum test z=3.43, p<0.005 and *slow* z=6.42, p<0.005). Finally, we even observed that the hand deviation was larger in the *slow* than in the *fast* condition (one-tailed ranksum test z=2.84, p<0.005).

Similarly, we observed that the final hand position along the x-axis, computed as the hand position when its velocity dropped below 2cm/s, exhibited similar dependency on the target condition. Indeed, a linear mixed model analysis (see section 2) revealed a significant effect of the target condition (β_1=-0.008$\pm$0.0003, t=-25.75, p<0.005). Similarly to what was observed for the maximal hand deviation, post-hoc pairwise analyses revealed that both dynamical conditions were characterized by less eccentric final hand positions than the no change condition (slow one-tailed ranksum tests, z=-11.18, p<0.005 and fast z=-17.72, p<0.005). These final hand positions in the dynamical conditions were more eccentric than the one in the switch condition (fast z=-5.884, p<0.005 and slow z=-12.85, p<0.005). Finally, the final hand positions in the slow condition were significantly more eccentric than those in the fast condition (z=-9.26, p<0.005). Trials that included a leftward mechanical perturbation contained the same effects: lateral hand deviation (linear mixed model: β_1=0.0033$\pm$0.0002, t=16.2, p<0.005) and final hand position along the x-axis (linear mixed model : β_1=0.008$\pm$0.0003, t=27.89, p<0.005) .

These differences observed in hand position suggest that the way the mechanical perturbations were handled differed in the different the target conditions. More specifically, it suggests that participants did not respond to the mechanical perturbations the same way in the dynamical and static target conditions and that they could even adjust their responses to the rate of the dynamical target condition. In order to investigate participants' responses to the mechanical perturbations

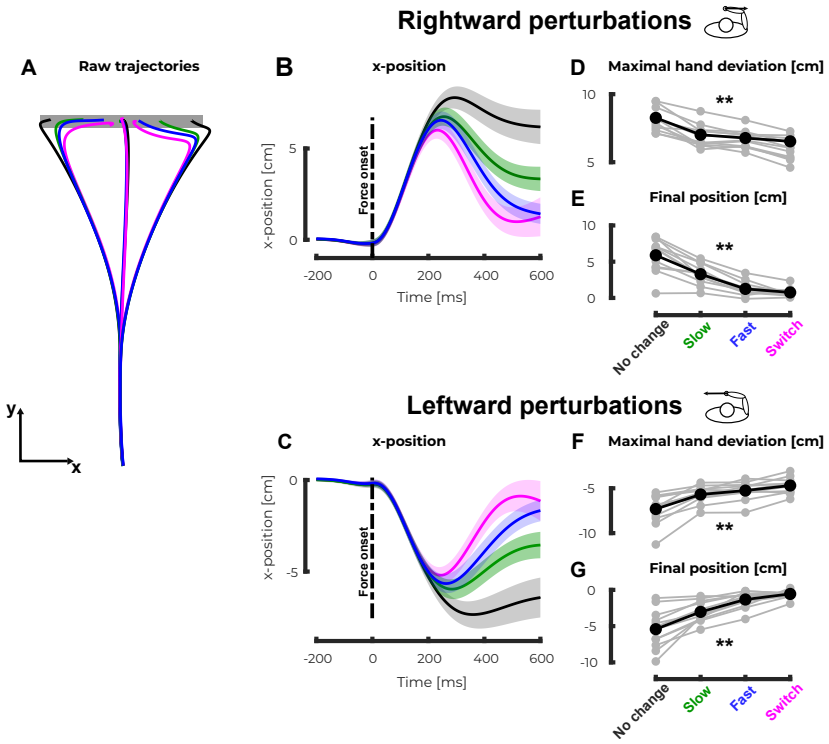

Figure III.2: Hand kinematics during movements. **A**. Group mean of the hand path for unperturbed and perturbed trials in the *no change* (black), *slow* (green), *fast* (blue) and *switch* (magenta) conditions. **B**. Group mean and SEM of the x-position of participants' hand s a function of time (aligned on force onset) for trials perturbed with rightward mechanical perturbations in the four target conditions. The black dashed line represents the onset of the mechanical perturbation. **C**. Group mean and SEM of the x-position of participants' hand as a function of time for trials perturbed with leftward mechanical perturbations in the four target conditions. The black dashed line represents the onset of the mechanical perturbation. **D**. Group mean (black) and individual means (gray) of the maximal hand deviation in presence of a rightward perturbation for the four target conditions. **E**. Group mean (black) and individual mean (gray) of the final position for trials with rightward perturbation for the four target conditions. **F**. Group mean (black) and individual means (gray) of the maximal hand deviation in presence of a leftward perturbation for the four target conditions. **G**. Group mean (black) and individual mean (gray) of the final position for trials with leftward perturbation for the four target conditions. ** $p<0.005$

we explored the lateral acceleration profiles. Since we wanted to study the responses specific to the change in target condition, we subtracted the acceleration profile in the no change condition from the other three conditions.

We observed that the lateral acceleration profiles were modified by the target condition. Indeed, Figures III.3A and B represent the target specific acceleration profiles obtained by subtracting the acceleration profile in the no change condition from each of the other ones for both perturbation directions. We observed that the first acceleration peak value after force onset, corresponding to participants' response to mechanical perturbation, depended on the target condition. A linear mixed model analysis (see section 2) performed on the value of this first acceleration peak revealed a significant effect of condition for both rightward (β_1=-0.24$\pm$0.017, t=-14.23, p<0.005) and leftward mechanical perturbations (β_1=-0.25$\pm$0.025, t=-10.08, p<0.005). These negative values for the β_1 estimate indicated that this peak value increased from *slow* to *switch* condition. Post-hoc analyses performed on pairwise comparisons confirmed this effect for both rightward and leftward perturbations (*fast* one-tailed ranksum tests : rightward z=-6.42, p <0.005 and leftward z=-2.95, p<0.005 and *slow* rightward z=-12.03, p<0.005 and leftward z=-8.85, p<0.005). Moreover, these pairwise comparisons also reported that this acceleration peak was larger in the *fast* than in the *slow* condition (one-tailed ranksum tests: rightward z=-6.28, p<0.005 and leftward z=-6.27, p<0.005).

3.1.2 Muscle activity

These kinematics results indicated that participants were able to adjust their control policy during movements to respond to dynamical changes in task constraints. They were even able to alter their adjustments to the rate of these dynamical changes. We hypothesized that the EMG activities in Pectoralis Major (PM) and Posterior Deltoid (PD), respectively the muscles stretched by the rightward and leftward perturbations, also depended on the target condition.

The target condition modulated the EMG activity of the muscles stretched by the mechanical perturbation. Figures III.4A and B represent the mean EMG activities collapsed across participants for trials perturbed by rightward or leftward perturbation in all target conditions in the stretched (full lines) and shortened muscles (dashed lines). Visual inspection of target specific responses for the stretched muscles, obtained by subtracting the *no change* condition to the other ones, confirmed this modulation of the EMG response (see FigureIII.4C and D respectively for PM and PD). In order to characterize this modulation, the

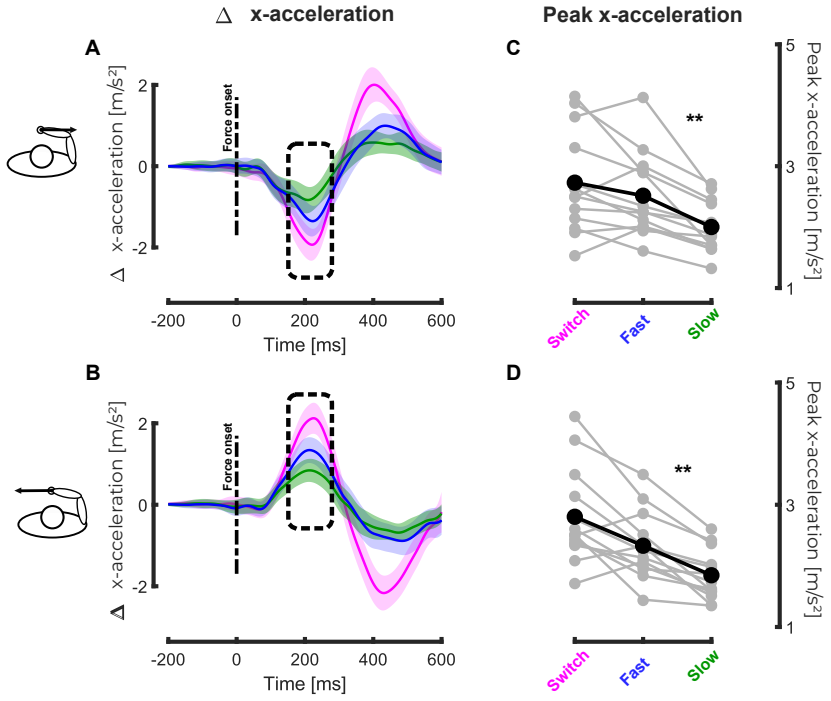

Figure III.3: Lateral acceleration during movements: **A** Group mean and SEM of target-specific lateral acceleration profiles for trials perturbed by rightward perturbation in the *switch* (magenta), *fast* (blue) and *slow* (green) target conditions, aligned on force onset. These profiles were computed by subtracting the acceleration profile in the *no change* condition. The black dashed line represents the mechanical perturbation force onset. **B** Group mean and SEM of target-specific lateral acceleration profiles for trials perturbed by leftward perturbation in the *switch* (magenta), *fast* (blue) and *slow* (green) target conditions, aligned on force onset. **C** Group mean (black) and individual means (gray) of the absolute value of the first peak lateral acceleration in presence of rightward mechanical perturbation in the *switch*, *fast* and *slow* target conditions. **D** Group mean (black) and individual means (gray) of the absolute value of the first peak lateral acceleration in presence of leftward mechanical perturbation in the *switch*, *fast* and *slow* target conditions. ** $p<0.005$

EMG activity of the stretched muscle was binned in the long-latency (LL 50-100ms after force onset) and voluntary epochs (VOL 100-180 ms after force onset) for each perturbation direction. The deviations from the mean activity in these two time bins are reported in Figures III.4E and F for stretched PM in the LL and VOL windows at population (black) and individual (gray) levels.

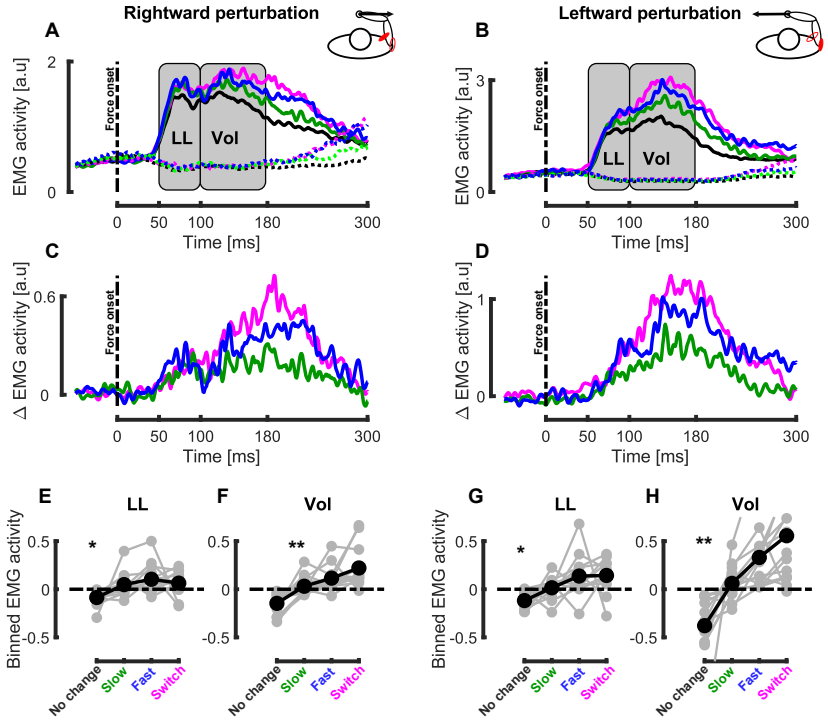

Figure III.4: EMG activity : **A** Group mean for the stretched (Pectoralis Major, full lines) and shortened (Posterior Deltoid, dashed lines) responses to rightward perturbations in the different target conditions. The gray rectangles represent the long latency (LL) and voluntary (VOL) epochs where the EMG activity was averaged to perform further statistical analyses. The black dashed line represents the mechanical perturbation onset and the time axis is aligned with force onset. **B** Group mean for the stretched (Posterior Deltoid, full lines) and shortened (Pectoralis Major, dashed lines) responses to leftward perturbations in the different target conditions. **C** Target-specific EMG responses to perturbations for Pectoralis Major in the presence of rightward perturbation for the *switch* (magenta), *fast* (blue) and *slow* (green) target conditions. Time axis is aligned with force onset. **D** Group mean of target-specific EMG responses to perturbation for Posterior Deltoid in the presence of leftward perturbation for the *switch* (magenta), *fast* (blue) and *slow* (green) target conditions. Time axis is aligned with force onset. **E-F** Group mean (black) and individual means (gray) of the binned EMG activity in the LL (**E**) and VOL (**F**) time windows for Pectoralis Major in the presence of rightward perturbations for the different target conditions. **G-H** Group mean (black) and individual means (gray) of the binned EMG activity in the LL (**G**) and VOL (**H**) time windows for Posterior Deltoid in the presence of leftward perturbations for the different target conditions. * p<0.05, ** p<0.005

We observed a significant effect of target condition on the modulation of the PM response in the LL (linear mixed model : β_1=-0.029$\pm$0.005, t=-5.70, p<0.005) and VOL window (linear mixed model : $\beta_1 = -0.060$

$\pm$ 0.00495, t=-12.24, p<0.005), respectively represented in Figures III.4E and F. Those negative values indicated larger responses for faster changes in target width. To further investigate these differences, we performed pairwise post-hoc comparisons between the different target conditions using linear mixed models (see section 2). In the LL window, we did not observe any difference between the different dynamical conditions (*switch*/*fast* : β_1=0.02$\pm$0.021, t=0.96, p=0.33, *switch*/*slow* $\beta_1 = -0.0038$
$\pm$0.013, t=-0.29, p=0.77 and *slow*/*fast* β_1=-0.0525$\pm$0.0419, t=-1.25, p=0.21), even though they all differed from the *no change* condition (p<0.005 for all conditions). However, these pairwise comparisons revealed significant differences in the VOL time window between the different dynamical conditions (*switch*/*fast* β_1=-0.05$\pm$0.019, t=-2.50, p=0.012, *switch*/*slow* β_1=-0.059$\pm$0.013, t=-4.35, p<0.005 and *slow*/*fast* β_1=-0.079$\pm$0.04, t=-1.96, p=0.048).

The same modulation of the EMG activity with the target condition was observed in PD for both LL (β_1=-0.046$\pm$0.007, t=-6.42, p<0.005 see FigureIII.4G) and VOL time epochs (β_1=-0.015$\pm$0.008, t=-18.66, p<0.005 see FigureIII.4F) as it was stretched by leftward perturbations. Interestingly, the pairwise post-hoc comparisons revealed significant differences between the dynamical conditions in both the LL (*switch*/*fast* : β_1=-0.005$\pm$0.019, t=-0.03, p=0.97, *switch*/*slow* β_1=-0.03$\pm$0.012, t=-2.336, p=0.019 and *slow*/*fast* β_1=-0.1193$\pm$0.057, t=-2.09, p=0.036) and the VOL time window (*switch*/*fast* : β_1=-0.072$\pm$0.021, t=-3.35, p<0.005, *switch*/*slow* β_1=-0.12$\pm$0.016, t=-7.33, p<0.005 and *slow*/*fast* β_1=-0.25$\pm$0.058, t=-4.33, p<0.005). These differences indicated that both reflexive and voluntary responses were modulated to the dynamical change in target width and that they were even tuned to the rate of change.

Altogether, these results indicate that participants adjusted their behavior during movements in response to dynamical changes in target shape. Indeed, we showed that the hand deviation induced by the mechanical perturbations was different in the dynamical (*slow* and *fast*) and in the static conditions (*no change* and *switch*). Moreover, we reported larger hand deviation for the *slow* than for the *fast* condition: indicating that the rate of change in target width was integrated in the control policy. The differences observed in acceleration profiles and EMG correlates confirmed this finding. The sensitivity of the online adjustments of

control policy to dynamical changes in target changes and the differences observed between the two different speeds suggest a dynamical an continuous re-evaluation of control policy within movement.

3.2 Differences between the first and last trials in dynamical conditions

Interestingly, we also observed differences in the hand kinematics between the first and the last trials in the *fast* and *slow* conditions. Figures III.5A and B represent the mean and SEM of the position along the x-axis for the first (full line) and last trials (dashed line) in the *fast* condition in presence of rightward and leftward mechanical perturbations respectively. We observed that these first and last trials differed and decided to take a look at their acceleration profiles in order to quantify these differences. The corresponding acceleration profiles are represented in Figures III.5C and D. We observed a consistent and significant lag of the last trial with respect to the first one. This lag was computed as the median of the maximal value of the cross-correlation performed with bootstrap resampling based on the fourteen subjects (see section 2). The resulting distribution of this lag, obtained through this resampling method is represented in Figure III.5E. This method revealed a median time lag of -18ms in the *fast* condition (depicted with a blue vertical line in Figure III.5) that was significantly smaller than zero (signrank test $z=-32.18$, $p<0.005$). No time lag was observed in the *slow* condition.

The first and last trials of each dynamical condition also differed in the smoothness of their acceleration profile as shown in Figures III.5C and D for the *fast* condition. This difference in smoothness was quantified by comparing the integral of the absolute values of the derivatives of these acceleration profiles: the jerk. We reported in Figures III.5F and G these integrals for all participants in the *slow* and *fast* conditions respectively. In the *fast* condition, the final state was less jerky than the first one as reported by a signrank test ($z=-2.835$, $p<0.005$). Similar results were obtained in the *slow* condition (signrank test $z=-3.19$, $p<0.005$) indicating an increase in the smoothness of the acceleration profiles.

These results indicated that, besides the online adjustments in control policies induced by dynamical changes in target width, participants also adapted their behavior across trials.

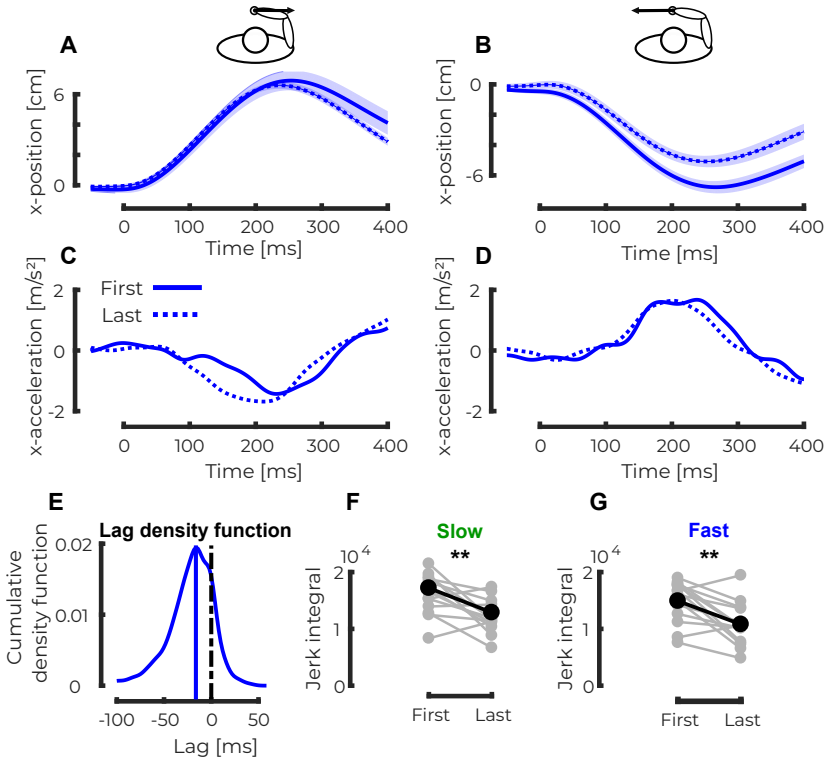

Figure III.5: Between trials analyses : **A** Group mean and SEM collapsed across participants of the first (full line) and last (dashed line) trials in the *fast* condition with rightward mechanical perturbation. Time axis is aligned on force onset. **B** Group mean and SEM collapsed across participants of the first (full line) and last (dashed line) trials in the *fast* condition with leftward mechanical perturbation. The black dashed line corresponds to the difference between the first and last trials. Time axis is aligned on force onset. **C** Group mean of the acceleration profiles of the first (full line) and last (dashed line) trials in the *fast* condition with rightward mechanical perturbation. Time axis is aligned with force onset. **D** Group mean of the acceleration profiles of the first (full line) and last (dashed line) trials in the *fast* condition with leftward mechanical perturbation. Time axis is aligned with force onset. **E** Cumulative density function of the lag between the last and first acceleration profiles for both perturbation directions. The blue vertical line corresponds to the median value. **F** Group mean (black) and individual means (gray) of the integral of the absolute value of the jerk for the first and last trials in the *slow* target condition. **G** Group mean (black) and individual means (gray) of the integral of the absolute value of the jerk for the first and last trials in the *fast* target condition. ** $p<0.005$

4 Discussion

We investigated whether humans can dynamically adjust their control policy used to perform reaching movements online if the task constraints were dynamically changed. More specifically, we studied participants' behaviour as they were reaching to a target whose width continuously decreased with time. We observed that the way participants responded to unexpected mechanical perturbations depended on the target condition and specifically on the rate of change in target width. Which indicated that the control policies used to perform movements were adjusted online to the specific change in target width. Interestingly, the EMG responses of the stretched muscles revealed that both the reflexive and voluntary responses to mechanical perturbations were sensitive to the specific rate of change of target width. Altogether, our results demonstrate that the control policies used by humans to perform reaching movements are dynamically adjusted within movements to respond to the changes in task constraints.

These findings expand the current view of goal-directed reaching movements by providing the evidence for dynamical adjustments of the control policy. Optimal Feedback Control (OFC) theory [Todorov and Jordan, 2002] assumes that reaching movements are executed by using control policies that are tuned to the task goal. This theory posits that the selected control policy minimises a cost-function quantifying the desired behavioural performance through a weighted sum of motor cost and motor errors [Todorov, 2005, Scott, 2004, Shadmehr and Krakauer, 2008]. Many task-related parameters such as the shape of the goal target [Knill et al., 2011, Nashed et al., 2012] or the reward associated with the task [Esteves et al., 2016, Summerside et al., 2018, Yoon et al., 2018] influence this cost-function and therefore its associated optimal control policy. This task-specific controller can handle unexpected perturbations that interfere with the performance of the task such as mechanical perturbations [Nashed et al., 2012, Lowrey et al., 2017] or target jumps [Georgopoulos et al., 1981, Franklin and Wolpert, 2008, Dimitriou et al., 2013], by providing state-feedback responses that are specifically tuned to the task performed. Here, we leveraged an experimental paradigm developed in previous work [De Comite et al., 2021], consisting in abrupt changes in target structure within movements, to dynamically change the task constraints and investigate whether

participants' control policy dynamically adjusted. The differences we observed between the two dynamical conditions and the condition were no change in target occurred indicate that dynamical changes in task constraints induce dynamical adjustments of the control policy. Moreover, the difference observed between the fast dynamical condition and the switch condition demonstrated that these control policies were dynamically adjusted during movement. We therefore confirm the existence of a mechanism whereby the control policy is dynamically adjusted during movement to be optimally tuned to the task performed.

The dynamical online adjustments of control policy could occur through different mechanisms. A first hypothesis would be that participants selected a control policy adapted to the rate of change in target width once they identified it. This suggest that participants somehow prepared multiple control policies. Many studies demonstrated that such parallel encoding of motor plans exists for multiple targets [Chapman et al., 2010, Gallivan et al., 2017, Wong et al., 2017] and even multiple target shapes [Gallivan et al., 2016b]. Our experimental paradigm differed from those used in these studies in that the multiple options were never presented at the same time and changes between targets occurred within movements. Moreover, the specific timings of our task (i.e. 150ms between the change in target condition and the beginning of a difference in the EMG activities) are such that this hypothesis is very unlikely. Indeed, we reported in a previous study that it takes approximately 150ms, from the onset of the visual change, to adjust the control policy online in response to change in target redundancy [De Comite et al., 2021], the additional differences observed between the dynamical and static conditions suggest that participants were able to estimate the rate of change as well. Previous reports demonstrated that a delay of about 100ms is required to be able to estimate the speed of a target in a smooth pursuit paradigm [Bahill and McDonald, 1983, Barnes and Asselman, 1991, Orban de Xivry et al., 2008]. The combination of these two time delays make it very unlikely that participants could switch between different preplanned motor policies. The alternative hypothesis would be that they used some adaptive policies consisting of successive small adjustments of control policies mediated through feedback mechanism.

The hypothesis whereby the online adjustments of control policies are handled through successive small adjustments implicitly assumes an overlap of movement planning and execution throughout the whole movement [Orban de Xivry et al., 2017]. This is an important result as it demonstrates that there is a dynamical feedback loop integrated in the reaching control policy that convey information relative to the task demands throughout the movement.

Between trials adjustment also occurred in our experiment even though they were small. Indeed, we could quantify that the movements were less jerky at the end of the experiment and that the response to perturbation in the fast condition were shifted forward in time. These between trials adjustments of behavior are reminiscent of the broad body of literature concerning motor learning [Shadmehr and Mussa-Ivaldi, 1994, Krakauer et al., 1999]. Albeit this parallelism, our paradigm remains very different from those investigating motor learning. Indeed, here we investigated how changes in task constraints happening within movement could influence the control policy while motor learning strive to understand how unmodeled environmental changes, such as force field [Krakauer et al., 1999, Lackner and Dizio, 1994, Shadmehr and Mussa-Ivaldi, 1994, Singh and Scott, 2003] and visuomotor rotation [Krakauer et al., 2000, Mazzoni and Krakaueur, 2006] for example, are handled by the control policy on a trial-to-trial basis. More recently, some motor learning studies reported within trials corrections in presence of velocity-dependent force fields [Crevecoeur et al., 2020b, Crevecoeur et al., 2020a, Mathew et al., 2020], but these are still different as we investigated online adjustments of the control policy induced by online modification of the task goal. To put that into the context of Optimal Feedback Control, in our study we probed participant's ability to adapt to online adjustment of the cost function while in motor learning studies they investigated participant's ability to adapt to unmodeled plant dynamics. The between trials adjustments we reported only mean that humans' ability to adjust their control policy online is getting better as they are getting more familiar with the task. However, whether the mechanisms underlying online adjustments in control policy and those underlying motor adaptation could be related is still unknown and could be a fascinating research question for further studies.

To summarize, we reported here that humans are able to dynamically adjust their control policy when they experience dynamical changes in task constraints. This finding highlights the existence of a feedback mechanism regulating the selection and application of control policy in human reaching movements. Moreover, this study also demonstrates that this feedback mechanism could be improved through habituation in order to give rise to smoother and more optimal movements.

Chapter IV

Reward-dependent selection of feedback gains impacts online motor decisions

You are the only responsible for your own wants

Isaac Asimov

Submitted as : De Comite A., Crevecoeur F., and Lefèvre P., Reward-dependent selection of feedback gains impacts rapid motor decisions, May 2021

Target reward influences motor planning strategies through modulation of movement vigor. Considering current theories of sensorimotor control suggesting that movement planning consists in selecting a goal-directed control strategy, we sought to investigate the influence of reward on feedback control. Here we explored this question in three human reaching experiments. First, we altered the explicit reward associated with the goal target and found an overall increase in feedback gains for higher target rewards, highlighted by larger velocities, feedback responses to external loads, and background muscle activity. Then, we investigated whether the differences in target rewards across multiple goals impacted rapid motor decisions during movement. We

observed idiosyncratic switching strategies dependent on both target rewards and, surprisingly, the feedback gains at perturbation onset: the more vigorous movements were less likely to switch to a new goal following perturbations. To gain further insight into a causal influence of the feedback gains on rapid motor decisions, we demonstrated that biasing the baseline activity and reflex gains by means of a background load evoked a larger proportion of target switches in the direction opposite to the background load associated with lower muscle activity. Together, our results demonstrate an impact of target reward on feedback control and highlight the competition between movement vigor and flexibility.

1 Introduction

From the toddler picking their favorite toys to the footballer selecting the best path through opponents, humans manifest the exquisite ability to plan and select movements. Movement planning is the process that integrates many task-related factors in order to select the best control strategy for the task (Wong et al., 2015). Amongst these numerous factors, the reward associated with the task induces a modulation of movement vigor in saccadic eye movements [Manohar et al., 2015, Manohar et al., 2017] and upper limb reaching movements [Summerside et al., 2018, Esteves et al., 2016]. Moreover, recent studies reported that higher reward increases visuomotor responses to disturbances [Carroll et al., 2019] and that the increase of vigor associated with reward is correlated with a reduction of movement variability and an increase in co-contraction [Codol et al., 2020]. Together, these previous results suggested an influence of reward on movement planning strategies.

Besides this impact on movement planning, reward also has an influence on movement selection. Indeed, the selection of the best alternative between different options is biased toward movements associated with the highest reward [Trommershäuser et al., 2003, Trommershäuser et al., 2008]. Similarly, when humans have to select a target, their choice is biased by parameters such as the biomechanical costs incurred when reaching to each potential option, resulting in target selection toward less effortful movements [Cos et al., 2011, Morel et al., 2017].

The commitment to an action actually results from a distributed consensus between low level sensorimotor representations of movement costs (e.g. motor costs) and high level cognitive representations of their outcomes (e.g. reward) [Cisek, 2012]. Here we explored the impact of target reward on fast feedback control strategies and tested the distributed consensus theory in a dynamical context by probing the effect of movement reward on feedback control and online motor decisions. Recent studies have sought to investigate whether and how much the factors that characterize action selection during movement planning could also influence movement selection when the hand has already started moving. A first body of work have shown that dynamical changes in target selection can be triggered by mechanical [Nashed et al., 2014] or visual [Kurtzer et al., 2020, Michalski et al., 2020] perturbations occurring during movement. More recently, some studies demonstrated that cognitive factors, such as the reward distribution of a redundant target, also influence online motor decisions [Cos et al., 2021, Martí-Marca et al., 2020]. They revealed that the reward distribution of a redundant target influences online motor decisions and suggested a link between the state of the limb (position and speed) at perturbation onset and the outcome of the decision. However, whether the reward of competing alternatives or the level of muscle activity could influence online motor decisions has not been explored yet.

In the present work, we addressed the relationship between target reward and feedback control as well as online motor decisions by applying perturbations while participants performed reaching movements toward one or several targets that differed explicitly by their associated rewards. We hypothesized that the influence of reward on movement planning was linked to the selection of feedback gains, which could impact one's ability to flexibly change target during movement. In a first experiment, we investigated the influence of reward on feedback control strategies. We then investigated the impact of reward on feedback control when participants had the opportunity to reach to different goals. The goal of the third experiment was to study the competition between feedback gains and the ability to flexibly change movement goal during movement.

We first reproduced previous findings of reward-related increase in velocity toward the target. Importantly, we uncovered that this modulation was associated with an increase in feedback gains and muscle

activity. In a second experiment, we observed that the difference in reward between alternative goals could bias online motor decisions and, interestingly, found out that the overall increase in movement vigor was negatively correlated with the potential selection of a new target. Our third experiment confirmed that biases in feedback gains induced experimentally were negatively correlated with the ability to switch goal during movement. These findings demonstrate that movement reward modulates both planning and feedback control, and involves the peripheral motor system through modulation of muscle co-contraction and reflex gains. Moreover, we highlight that this modulation was detrimental to the ability to flexibly switch to a new goal during movement.

2 Methods

2.1 Participants

A total of 53 participants were enrolled in this study and took part to one of the three experiments. The first group performed Experiment 1 and included 14 right-handed participants (7 females) ranging in age from 21 to 27. The second group performed Experiment 2 and included 20 right-handed participants (14 females) ranging in age from 20 to 46. The last group performed Experiment 3 and included 19 right-handed participants (11 females) ranging in age from 18 to 52. Participants were naïve to the purpose of the experiments and had no known neurological disorder. The ethics committee of the host institution approved the experimental procedures and participants provided their written informed consent prior to the experiment.

2.2 Setup

For the three experiments, participants sat on an adjustable chair in front of a Kinarm end-point robotic device (KINARM, Kingston, ON, Canada) and grasped the handle of the right robotic arm with their right hand. The robotic arm allowed movements in the horizontal plane and direct vision of both the hand and the robotic arm was blocked. Participants were seated such that at rest their arm was vertical and their elbow formed an angle of approximately 90° degrees. Their arm was unconstrained and their forehead rested on a soft cushion attached to the frame of the setup. A virtual reality display placed above the

handle allowed the participants to interact with virtual targets. A white dot of 0.5 cm radius corresponding to the position of the handle was shown on this display during the whole experiment.

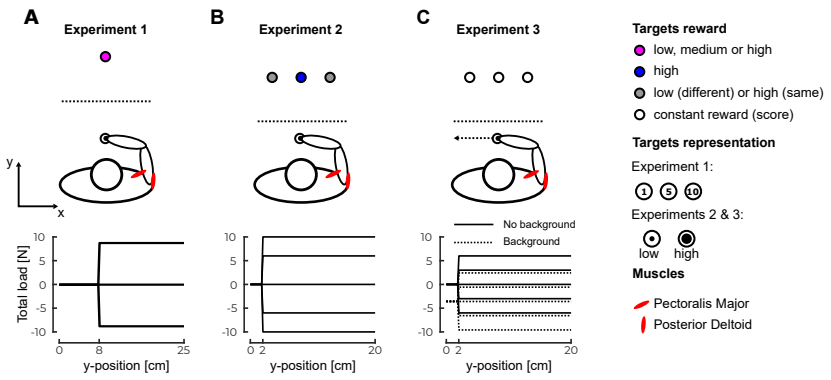

Figure IV.1: Task paradigms –. **A** Representation of the task paradigm of Experiment 1. Participants controlled a hand-aligned cursor represented by the black dot on a virtual reality display. They had to reach for the goal target, represented by the purple goal target in front of them. This goal target could have a low, medium or high reward (1, 5 or 10 points). The bottom part of the panel represents the load profiles that participants could experience. **B** Representation of the task paradigm of Experiment 2. Participants had to reach for any of the three targets presented in front of them. The central target always had a high reward whereas the two others either had a low or a high reward. The bottom part of the panel represents the perturbation load that participants could encounter during movements. **C** Representation of the task paradigm of Experiment 3. Participants had to reach for any of the three targets presented in front of them. During the second half of the trials a background load force directed leftward was applied prior and during the movement (dashed line bottom panel). The bottom part of the panel represents the possible profiles of the total load forces (perturbation load + background load). EMG data from Pectoralis Major and Posterior Deltoid were collected during all Experiments.

2.3 Experiment 1

In Experiment 1 (Figure IV.1A, top), participants (N=14) were instructed to perform reaching movements to a small circular goal target (1.5 cm radius) located at 25 cm in the y-direction from the start position, a red disk of 1.5 cm radius. Participants had first to put the hand-aligned cursor in the start position, which turned green as they reached it. After a random time delay (drawn from an uniform distribution between 1 and 2s), the goal target appeared as a red disk containing a number (1, 5 or 10) that corresponded to the reward participants would receive if they reached and stabilised within the target for a prescribed time

window. Reaction time was not constrained and participants could start the movement whenever they wanted. Following the exit of the start position, participants had up to 600ms to reach the goal target and keep the cursor inside for at least 500ms. The goal target turned green at the end of successful trials, or remained red otherwise. During movements, a mechanical perturbation load could be applied to participants' hand (33% of the trials). This load consisted of a lateral step force of ± 9 N, with a 10ms linear build-up, aligned with the x-axis. This force was triggered when the hand-aligned cursor crossed a virtual line located at 8 cm from the center of the start position (Figure IV.1A, bottom). Unperturbed and perturbed trials as well as trials with different rewards were randomly interleaved such that participants could not predict the occurrence or the direction of the perturbations. Participants started with a 27-trials training block in order to become familiar with the task and the force intensity of perturbation loads. After completing this training block, they performed 6 blocks of 72 trials interleaved with pauses of three to five minutes to prevent muscle fatigue. Each 72-trials block included: 48 unperturbed trials (16 with each target reward) and 24 trials which contained mechanical perturbations (leftward or rightward, 8 of each reward condition). Participants performed a total of 432 trials, including 24 for each perturbed condition (direction of the mechanical perturbation and value of the target reward). A total score corresponding to the cumulative sum of individual movement rewards was projected next to the goal target. Participants were compensated for their participation according to a conversion of this total score. This conversion was calculated such that each participant received between 10 and 15 € as an incentive to score a maximum number of points during the experiment.

2.4 Experiment 2

Experiment 2 was designed to assess the effect of reward on online motor decisions between competing motor goals. Instead of reaching to a single target, participants (N=20) were instructed to perform reaching movements to any of three circular targets (1.5 cm radius) located at 20 cm in the y-direction from the same start position as in Experiment 1 (Figure IV.1B, top). As in Experiment 1, the goal targets appeared after participants stabilised the hand-aligned cursor in the start position. All three goal targets appeared in each trial, the central one being aligned

along the y-axis with the start position and the other two equidistant from this central target at 9 cm along the x-axis. These targets were presented as an inner disk of radius 0.7 or 1.2 cm inside an outer circle of radius 1.5 cm. The purpose of the inner disk was to show the reward associated with the target: the larger the diameter of this disk, the higher the reward. There were two different conditions of reward: either all the targets had the same large reward (same values condition) or only the central target had a large reward while the other two had lower rewards (different values condition). In a pilot study, we considered a third reward configuration: the central target had a small reward while the other two had higher rewards. We observed that, in this third configuration, the behavior in the absence of perturbation load was biased toward the lateral targets. We therefore decided to exclude this condition to keep the conditions in which participants spontaneously reached for the center target for the largest proportion of trials in the absence of any perturbation load. After a random time delay (drawn from a uniform distribution between 1.5 and 3s), the inner disks of the goal targets turned white and participants had to reach any of these within 400ms to 1000ms to pass the trial. Similar to Experiment 1, the reaction time was not constrained. The trial was successfully completed if participants reached any goal target in the prescribed time window and stabilised the cursor in it for 500ms. The inner disks of the goal targets turned green if the trial was successful and red otherwise. As in Experiment 1, a mechanical perturbation load could be applied to participant's hand (50% of the trials, ± 6 N or ± 10 N, 10 ms build-up aligned with the x-axis Figure IV.1B, bottom). This perturbation was triggered when the hand-aligned cursor crossed a virtual line located at 2 cm from the start position. Unperturbed and perturbed trials as well as trials with different reward distributions and force intensities were randomly interleaved. Participants started with a 58-trials training block followed by 6 blocks of 80 trials. Pauses of three to five minutes were introduced between blocks to prevent muscle fatigue. Each 80-trials block included: 40 unperturbed trials and 40 trials which contained mechanical perturbations. Participants performed a total of 480 trials including 30 trials of each perturbation condition (reward condition and mechanical perturbation condition). Participants were compensated for their participation using the same conversion rule as in Experiment 1.

2.5 Experiment 3

The third experiment was a variant of Experiment 2 and was designed to test the possible impact of co-contraction on online motor decisions by applying a background force laterally to the reach path (Figure IV.1C, top). Participants had to perform reaching movements to any of the three targets located as in Experiment 2. These targets were identical to the large reward target of Experiment 2 and the time course of events in the trial was similar as well except that a leftward background mechanical load of 4N was applied as volunteers reached the home target and remained on throughout the trials. As in the Experiment 2, a mechanical perturbation load could be applied to participant's hand during movement (33% of the trials). This load consisted of a ± 3 N or ± 6 N with a 10ms build-up triggered when the hand-aligned cursor crossed a line located at 2 cm from the home target (Figure IV.1C, bottom) and was added to the background load. Participants first performed a 21-trials training block which did not involve background load. After completing this training, participants performed 4 blocks of 60 trials which did not included the background load. Each 60-trials block included: 40 unperturbed trials and 20 trials which included mechanical perturbations and they were interleaved with pauses of $3-5$min. After these 60-trials blocks, participants performed a second 21-trials training block which included the background load. Once this second training block was completed, participants performed a second set of 4 blocks of 60 trials which included the background load. They thus performed a total of 480 trials amongst which 24 of each condition (with different perturbation loads and background load on or off). To motivate participants, a score corresponding to their number of successful trials was projected next to the goal targets. Participants were compensated a fixed amount for their participation.

The third experiment was a variant of Experiment 2 and was designed to test the possible impact of muscle activity on online motor decisions by applying a background force orthogonally to the reach path (Figure IV.1C, top). Participants had to perform reaching movements to any of the three targets, located as in Experiment 2. These targets were identical to the large reward target of Experiment 2 and the time course of events in the trial was similar as well except that a leftward background mechanical load of 4N was applied as participants reached the start position and remained on throughout the trials. As in the Experi-

ment 2, a mechanical perturbation load could be applied to participant's hand during movement (33% of the trials). This load consisted of a ± 3 N or ± 6 N with a 10ms build-up triggered when the hand-aligned cursor crossed a line located at 2 cm from the start position (Figure IV.1C, bottom) and was added to the background load. Participants first performed a 21-trials training block which did not involve background load. After completing this training, participants performed 4 blocks of 60 trials which did not include the background load. Each 60-trials block included 40 unperturbed trials and 20 trials with mechanical perturbations and they were interleaved with pauses of $3-5$min. After these 60-trials blocks, participants performed a second 21-trials training block which included the background load. Once this second training block was completed, participants performed a second set of 4 blocks of 60 trials which included the background load. They thus performed a total of 480 trials amongst which 24 of each condition (with different perturbation loads and background load on or off). To motivate participants, a score corresponding to their number of successful trials was projected next to the goal targets. Participants were compensated a fixed amount for their participation.

2.6 Data collection and analysis

Raw kinematics data was sampled at 1kHz and low-pass filtered using a 4th order double-pass Butterworth filter with cut-off frequency of 20 Hz. Hand velocity along the y-axis was computed from numerical differentiation of the position data using a 4^{th} order centered finite difference.

Surface EMG electrodes (Bagnoli surface EMG sensor, Delsys INC. Natick, MA, USA) were used to record muscles activity during movements. We measured the Pectoralis Major (PM) and the Posterior Deltoid (PD) based on previous studies [Crevecoeur et al., 2019, De Comite et al., 2021] that showed that in this configuration they are stretched by the application of forces opposite to their action, and therefore largely recruited by the feedback responses. Before applying the electrodes, the skin of the participant was cleaned and abraded with cotton wool and alcohol. Conduction gel was applied on the electrodes to improve the quality of the signals. The EMG data were sampled at a frequency of 1kHz and amplified by a factor of 1000. A reference electrode was attached to the right ankle of the partici-

pant. Raw EMG data from the pectoralis major (PM) and posterior deltoid (PD) were band-pass filtered using a 4^{th} order double-pass Butterworth filter (cut-offs: 20 and 250 Hz), rectified, aligned to force onset and averaged across trials or time windows as specified in the Results section. The time windows selected for the temporal averaging are the short-latency (20-50ms), the long-latency (50-100ms) and the voluntary time epochs (100-180ms) as proposed in previous work [Pruszynski et al., 2008, Pruszynski and Scott, 2012].

EMG data were normalized for each participant to the average activity collected when participants maintained postural control at the start position against a constant force of 9 N. Data from the pectoralis major were normalized by the EMG activity in the same muscle while performing postural control against a rightward force whereas data from the posterior deltoid were normalized by the EMG activity in the same muscle while performing postural control against a leftward force. This calibration procedure was applied after the second and the fourth blocks in the first two experiments and after the first, third, fifth and seventh blocks in the third experiment. Data processing and parameters extractions were performed using MATLAB 2019a.

In Experiment 1, we fitted linear mixed models to determine the effect of the target reward on the kinematics and EMG activity. These models were fitted using the *fitlme* function of *Matlab* and the formula used was the following equation:

$$\text{Parameter} = \beta_0 + \beta_1 * \text{Reward} + \alpha_i * \text{Subject}_i \qquad \text{(IV.1)}$$

The fixed predictors were the intercept (β_0) and the reward condition (β_1) while the participants were included as a random offset (α_1). For all linear mixed model analyses that we performed, we reported the estimate for β_1, the t-statistics for this estimate as well as the corresponding p-value and the r^2 of the model. One tailed paired t-tests were used for post-hoc analyses where we collapsed across trials and participants to compare the different conditions. Effect size for these tests were reported using Cohen's d defined as the difference between the means of the two populations divided by the standard deviation of the whole sample.

To analyse the data from Experiments 2 and 3, we designed a multilinear logistic regression model to infer the effect of reward distribution and background load on target choice as the dependent variable, respectively. Considering that the dependent variable was a discrete variable (the chosen target), we use the following logistic regression model:

$$\log\left(\frac{P\,(\text{Lateral target})}{P\,(\text{Central target})}\right) = \beta_0 + \beta_1 * \text{Parameter}_1 + \beta_2 * \text{Parameter}_2 \tag{IV.2}$$

where the first effect (β_1) was the reward condition (Experiment 2) or the presence of a background load (Experiment 3) and the second effect (β_2) was the intensity of the perturbation load. For these logistic regressions, we reported the estimates for β_1 and β_2, their corresponding t-statistics as well as their p-value. For post-hoc analyses in Experiment 2, we used a one tailed Wilcoxon signed ranked tests for which we reported the ranksum, the z-statistics when provided, the p-value as well as the effect size given by the Cohen's d as defined above. In order to investigate the asymmetry in the parameters β_1 obtained in Experiment 3, we used bootstrap resampling on the individual data to generate 1000 estimates of the β_1 parameter for each condition (leftward perturbation versus rightward perturbation) using the multilinear logistic regression described above. We then assessed the asymmetry of the effect by investigating whether the 95% confidence interval of the difference between these two β_1 parameters contains 0 [Efron, 1979].

In order to determine the effect of the background load on the baseline muscle activity in Experiment 3, we fitted a linear mixed models with interaction terms following this equation:

$$\begin{aligned}\text{EMG} =& \beta_1 * \text{Background} + \beta_2 * \text{Muscle} + \\ & \beta_{12} * \text{Muscle} : \text{Background} + \alpha_i * \text{Subject}_i\end{aligned} \tag{IV.3}$$

Where the first term (β_1) refers to the background condition, the second (β_2) to the muscle, the third one (β_{12}) to the interaction term and the last one (α_i) to the random effect of participants. For all these β's, we reported their estimated value as well as their t-statistics, associated p-value and the r^2 of the model. Significance was considered at the level of $p = 0.05$ even though we decide to exactly report

any p-value that was larger than $p = 0.005$ as previously proposed [Benjamin and Berger, 2018]. In the figures, we reported significant differences for the level $p < 0.05$ (*), $p < 0.01$(**) and $p < 0.005$(***).

3 Results

3.1 Influence of the target reward on feedback corrections during movement

To determine whether target reward influences feedback corrections during movement, participants were instructed to perform reaching movement to a goal target associated with a reward that could change across trials (see section 2). During movements, mechanical perturbation loads could be applied to reveal feedback corrections. The occurrence of feedback corrections was assessed by looking at feedback kinematics and EMG responses of the muscles stretched by the perturbations.

The mean hand path trajectories across participants are represented in Figure IV.2A for the different perturbations and reward conditions. Consistent with previous work [Shadmehr et al., 2016, Summerside et al., 2018] we observed a significant increase in forward peak velocity with increasing reward value. Figure IV.2B represents the differences in the forward velocities between the high (dash-dot lines) or medium (full lines), and low reward conditions unperturbed (top) and perturbed (bottom) trials. The peak forward velocity increased with increasing reward value both for unperturbed (Figure IV.2D top, linear mixed model: $\beta_1 = 0.013$, $t = 6.51$, $p < 0.005$, $r^2 = 0.76$) and perturbed (Figure IV.2D bottom, linear mixed model, right: $\beta_1 = 0.014$, $t = 3.76$, $p < 0.005$, $r^2 = 0.789$, left: $\beta_1 = 0.018$, $t = 4.68$, $p < 0.005$, $r^2 = 0.79$) trials. Post-hoc comparisons between low and high reward conditions revealed a significant increase of peak velocity with reward for all perturbation conditions (one tailed paired t-tests, unperturbed: $t = -7.48$, $p < 0.005$, $d = 0.12$, left: $t = -5.37$, $p < 0.005$, $d = 0.16$ and right: $t = -3.99$, $p < 0.005$, $d = 0.13$). We did not observe any modulation of the reaction time required to initiate movement with the reward (linear mixed model, $p > 0.05$) since we instructed participants to initiate movement whenever they wanted.

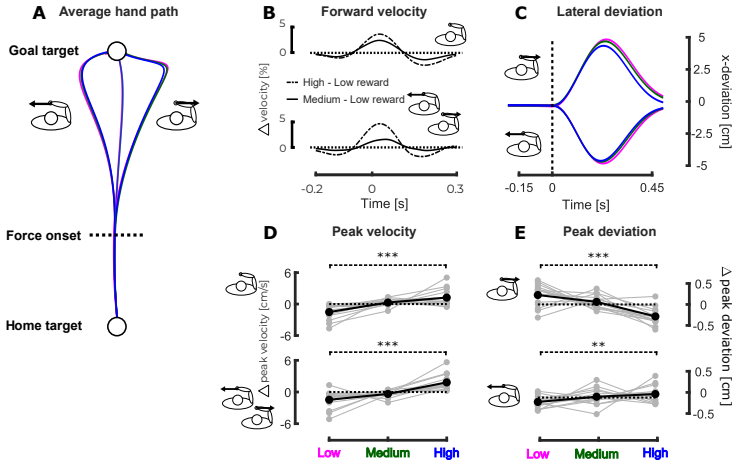

Figure IV.2: Experiment 1, kinematics - **A** Mean hand path across participants for the different conditions of Experiment 1. The magenta, green and blue traces respectively correspond to the low, medium and high reward conditions. The dashed line represents the onset of mechanical perturbations. **B** Mean differences in forward velocity between the high and low (dash-dot line) and medium and low (full line) for he unperturbed (top) and perturbed (bottom) trials. The time axis is aligned on force onset. **C** Mean hand path deviation across participants for the perturbed trials. The hand deviation has been obtained by subtracting the mean hand path to the perturbed hand path in the same reward condition for every subjects. The top part of the graph represents the trials perturbed to the right whereas the bottom part represents the trials perturbed to the left. **D** Group mean (black) and individual means (gray) of the differential forward peak velocity for the unperturbed trials (top) and perturbed trials (bottom) as a function of the reward condition with respect to average forward peak velocity. **E** Group mean (black) and individual means (gray) of the difference in hand deviation with respect to the mean hand deviation for leftward (top) and rightward (bottom) perturbations in the three reward conditions with respect to the average hand deviation. $p < 0.05$ (*), $p < 0.01$(**), $p < 0.005$ (***)

The effect of the mechanical perturbation on the movement kinematics was also dependent on the reward value. Indeed, the lateral hand deviation induced by the mechanical perturbation (Figure IV.2C), computed as the difference between the hand paths in the perturbed conditions and the mean hand path in the corresponding unperturbed reward condition for each participant, was significantly dependent on the reward condition. For both perturbation directions (Figure IV.2E top for leftward and bottom for rightward perturbations), we observed a significant decrease in the maximal hand deviation along the x-axis with increasing reward value (linear mixed models, right: $\beta_1 = -0.0025$, $t = -4.98$, $p < 0.005$, r^2 =0.38 and left : $\beta_1 = -0.0009$, $t = -2.25$,

$p = 0.024$, $r^2 = 0.35$). Post-hoc comparisons between low and high reward conditions revealed a significant decrease for both perturbation directions (one-tailed paired t-tests, right: $t = 5.31$, $p < 0.005$, $d = 0.14$ and left: $t = 2.34$, $p = 0.009$, $d = 0.3$).

Based on these kinematics analyses and previous studies showing that faster movements and small hand deviations induced by perturbations are correlated with high EMG activity [Crevecoeur et al., 2019] we hypothesised that the EMG activity in PM and PD during movement scaled with increasing reward. We investigated this effect both for baseline activity measured during unperturbed trials and for feedback responses to perturbation loads.

We observed a positive relationship between the EMG activity during unperturbed trials and the value of the target reward. Figure IV.3A represents the mean EMG activity collapsed across muscles and participants for unperturbed trials while the differences between these collapsed EMG activities in the high (dash-dot line) or medium (full line) and the low reward condition are represented in Figure IV.3B. We binned the EMG activity of each trial in a time bin ranging from 0 to 200ms after perturbation onset (gray rectangle in Figure IV.3A and Figure IV.3B) and fitted a linear mixed model (see section 2) on these binned values to determine whether reward had an influence on the EMG activity (deviations from the mean binned EMG activity in the different reward conditions are represented in Figure IV.3C and Figure IV.3D for PM and PD respectively). We observed an increase in EMG activity with the reward in both muscles (PM: $\beta_1 = 0.028$, $t = 4.603$, $p < 0.005$, $r^2 = 0.68$, PD: $\beta_1 = 0.053$, $t = 5.98$, $p < 0.005$, $r^2 = 0.66$). Post-hoc analyses performed on individual data showed that EMG activity was larger in the high reward condition than in the low one for both muscles (one tailed paired t-tests: pectoralis, t=4.14, $p < 0.005$, $d = -0.118$ and deltoid, $t = -2.92$, $p = 0.0059$, $d = 0.1653$).

The EMG response to mechanical perturbation in the agonist muscles was also modulated by the reward value. Indeed, linear mixed model analyses performed on the responses measured in PM and PD, when respectively a rightward or leftward perturbation occurred, showed a significant increase of EMG activity with increasing reward in the long-latency epoch ($50 - 100$ ms). We reported the EMG activities collapsed across muscles and participants in Figure IV.3E as well as the difference in these activities between the high (dash-dot line) or

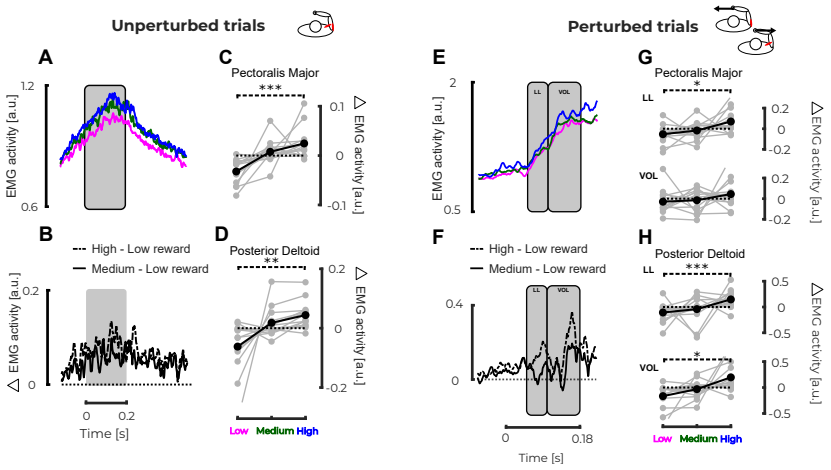

Figure IV.3: Experiment 1, EMG activity - **A** Mean EMG activity collapsed across muscles and participants for unperturbed trials. The time axis is aligned on force onset. **B** Mean differences in EMG activity collapsed across muscles and participants between high and low (dash-dot line) and medium and low (full line) reward conditions for unperturbed trials. **C** Group mean (black) and individual means (gray) of Pectoralis Major EMG activity binned between 0 and 200ms after force onset for unperturbed trials. **D** Group mean (black) and individual means (gray) of Posterior Deltoid EMG activity binned between 0 and 200ms after force onset for unperturbed trials. **E** Mean EMG activity collapsed across muscles and participants while responding to perturbations. Only the activities of the agonist muscles (ie. PM for rightward and PD for leftward perturbations) were used. **F** Mean differences in EMG activity collapsed across muscles and participants between high and low (dash-dot line) and medium and low (full line) reward conditions for agonist muscles in presence of perturbation load. **G** Group mean (black) and individual means (gray) differential EMG activity in PM binned in the long latency (50-100ms, top) and voluntary epochs (100-180ms, bottom) as a function of the reward condition. **H** Group mean (black) and individual means (gray) of the differential EMG activity in PD binned in the long latency (50-100ms, top) and voluntary epochs (100-180ms, bottom) as a function of reward condition. $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.005$ (***)

medium (full line) and the low reward condition in Figure IV.3F. For each perturbation direction, we binned the EMG activity of the stretched muscle in the long latency (LL $50 - 100$ ms after force onset) and voluntary (VOL $100 - 180$ ms after force onset) epochs. Figure IV.3G and Figure IV.3H respectively represent the deviation from the mean binned EMG activity in these two time bins (LL top and VOL bottom) for PM and PD in the different reward conditions. In PM we observed a significant increase in LL window (mixed model: $\beta_1 = 0.0615$, $t = 2.89$, $p < 0.005$, $r^2 = 0.64$) but even though a positive tendency emerges in VOL window, no significant increase was observed (mixed model: $\beta_1 = 0.036$, $t = 1.616$, $p = 0.106$, $r^2 = 0.69$). Individual pairwise post-hoc comparisons between low and high conditions confirmed these findings (one tail paired t-tests: LL, $t = -2.48$, $p = 0.0137$, $d = 0.1592$ and VOL, $t = -1.18$, $p = 0.128$, $d = 0.08$). The same holds for PD in which we found a significant increase of EMG activity in LL window with the reward (mixed model: $\beta_1 = 0.216$, $t = 2.12$, $p = 0.034$, $r^2 = 0.63$) but no significant effect in the VOL window (mixed model: $\beta_1 = 0.181$, $t = 1.887$, $p = 0.059$, $r^2 = 0.77$). In this case however, the individual pairwise comparisons between low and high conditions revealed a significant increase in both time windows (one tailed paired t-tests: LL, $t = -3.68$, $p < 0.005$,$d = 0.11$ and Vol, $t = -2.57$, $p < 0.07$, $d = 0.0116$).

An interesting question is whether the effect of target reward reported here could only be attributable to higher movement speed. In other words, could it be that the impact of a higher reward is an increase in movement speed that would therefore modulate the behaviour. To answer that question, we compared the lateral hand deviation observed in trials that have similar peak velocity and investigated whether, in these trials, the reward condition modulates the lateral hand deviation. We performed a linear mixed model analysis on the absolute values of these lateral hand deviations and reported an effect of the reward condition: smaller deviations for higher reward value ($\beta_1 = -0.0011$, $t = -2.61$ and $p = 0.009$). This result confirms that reward does not only modulate movement vigor but also the feedback responses to mechanical perturbations.

Therefore, the results of Experiment 1 revealed that the value of the target reward influenced both the movement kinematics and the EMG activity recorded during movement. Indeed, we showed that the hand

deviation induced by mechanical perturbation decreases with increasing reward value for both rightward and leftward perturbations. Moreover, the forward peak velocity of reaching movement increased with increasing reward value. Finally, EMG activity in both PM and PD increased with increasing reward value for unperturbed trials and in the long-latency response window for perturbed movement when the muscles were stretched by the perturbation. The modulation of forward hand velocity and baseline EMG activity which also produces increases in feedback responses to perturbation loads is consistent with an increase in control gains previously observed in uncertain dynamical contexts and interpreted as a robust control policy [Crevecoeur et al., 2019].

3.2 Influence of the reward of the potential options on online motor decisions

In Experiment 1, we showed that reward of the goal target had an influence on the way humans perform reaching movements to this target similarly to other task parameters such as target shape, presence of obstacles... Moreover, previous studies have shown that these task parameters that modify the control policies could also influence online motor decisions [Nashed et al., 2014]. We therefore designed a second experiment to determine whether reward could also influence online motor decisions. In this second experiment, participants had to reach to any of three potential targets aligned orthogonally to the main reaching direction (see section 2). The central target always had a large reward while the two lateral targets could either have lower reward or a reward equal to that of the central target. We assessed the effect of the difference between central and lateral rewards on online motor decisions by investigating the frequencies of reaching for the lateral targets. Mechanical perturbations that could occur during movement were used to evoke change in goal target. Because perturbations were unpredictable, a change in reaching frequency for the lateral targets dependent on the perturbation load was indicative of a perturbation-mediated change in goal that occurred during movement. The biomechanical and EMG states at perturbation onset in the same condition were also investigated to determine whether they had an influence on the future decision.

First, we observed that the reward of the lateral targets had a clear effect on the frequency of trials that ended on these targets. Figure IV.4A represents the hand paths of a typical participant toward the different

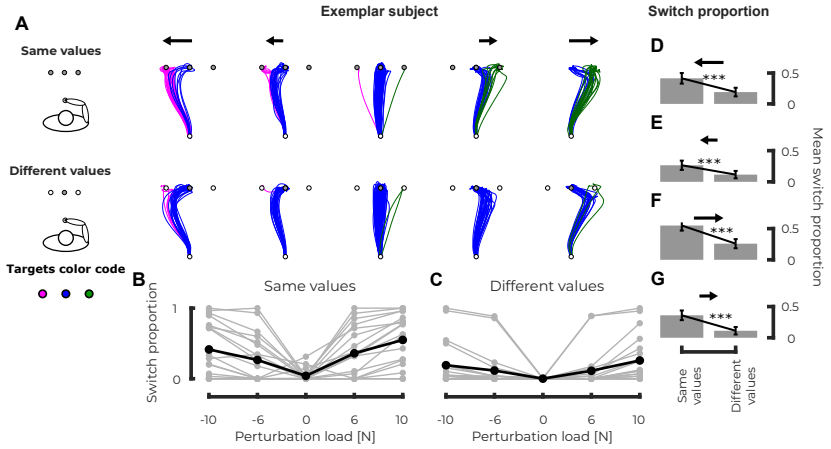

Figure IV.4: Experiment 2, Kinematics - **A** Representation of hand path of individual trials for a representative subject in the *same values* condition on top (all the targets had the same reward) and in the *different values* condition on bottom (the central target had a higher reward than the other two). The different columns represent the different force levels (from right to left, large leftward perturbation to large rightward perturbation). Magenta, blue and green paths respectively represent the trials that reached the left, central and right targets. **B** Group mean (black) and individual means (gray) of the switch proportion (i.e. fraction of trials that reached either the left or right target) as a function of the applied load for the *same values* condition. **C** Group mean (black) and individual means (gray) of the switch proportion (i.e. fraction of trials that reached either the left or the right target) as a function of the applied load for the different conditions. **D-G** Comparison of the switch proportion for the same (left) and different (right) conditions for the trials with large leftward force, small leftward force, small rightward force and large rightward force respectively. $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.005$ (***)

targets in various conditions. In general, in the absence of perturbation, people tend to reach to the central target except for some trials ($< 1\%$ in the *different values* condition and 8% in the *same values* condition). In all cases, the frequency of reach to the lateral targets increased with the magnitude of the perturbation (top and bottom part of Figure IV.4A for *same* and *different values* conditions respectively). In addition, there was a significant effect of the lateral target rewards on the frequency of lateral target: lower frequencies for different rewards. In order to determine the significance of these effects, we identified for each trial the target that was reached at the end of the movement and fitted a multilinear logistic regression on these data to determine whether the reward condition and the force had an influence on the target reached (see section 2). We observed a significant effect of the reward for both

the left ($\beta_1 = 1.103$,$t = 10.84$,$p < 0.005$) and right ($\beta_1 = 1.666$,$t = 18.45$,$p < 0.005$) targets versus the central one. These positive values indicate that the reach proportion to the lateral targets is larger in the same than in the different conditions (see Figure IV.4B and Figure IV.4C for the *same* and *different values* conditions, respectively). Intensity of the perturbation loads also had a significant effect for both lateral targets versus the central one (left: $\beta_1 = -1.33$, $t = -26.82$, $p < 0.005$ and right: $\beta_1 = 1.25$, $t = 29.07$, $p < 0.005$). Due to the sign of the force kept in the regression model, in both cases the frequency of lateral target reached increased with the force magnitude in absolute value. Post-hoc analyses performed at fixed force levels showed significant effect of the reward condition on the reaching proportion for all the perturbed conditions. We observed a smaller reach proportion to the left target in the *different values* condition compared to the *same value* condition for both perturbation directions (one tailed Wilcoxon signed-rank test: ranksum $= 3$, $p < 0.005$, $d = 0.61$ - see Figure IV.4D - and ranksum $= 1$, $p < 0.005$, $d = 0.49$ – see Figure IV.4E - for loads of -10 and -6N respectively). The mirror effect was observed for the right target: a decrease in the reach proportion in the *different values* condition for both perturbation directions (one tailed Wilcoxon signed-rank test: ranksum $= 0$, $p < 0.005$, $d = 0.74$ – see Figure IV.4F - and ranksum $= 3$, $p < 0.005$, $d = 0.78$ – see Figure IV.4G - for loads of 6 and 10 N respectively).

These results showed that participants took the reward distribution of the options offered by the three targets into account while deciding which target they should reach. The next question that we will address is whether any parameters linked to the current state of the limb could modify the decision between the different motor outcomes.

Interestingly, we observed a link between the state of the limb at perturbation onset (kinematics and EMG activity) and the outcome of the motor decision. Figure IV.5A represents the mean EMG activity recorded in PM (top) and PD (bottom) in presence of mechanical perturbations (rightward, first column and leftward second column) across participants for the different targets (magenta: left, blue: centre and green: right) in the *same values* condition. No significant differences were observed in PM prior to force onset (-150ms to 0ms, grey rectangle in Figure IV.5A) between the trials that reached the center target and the ones that reached the lateral targets (Figure IV.5B top) for both

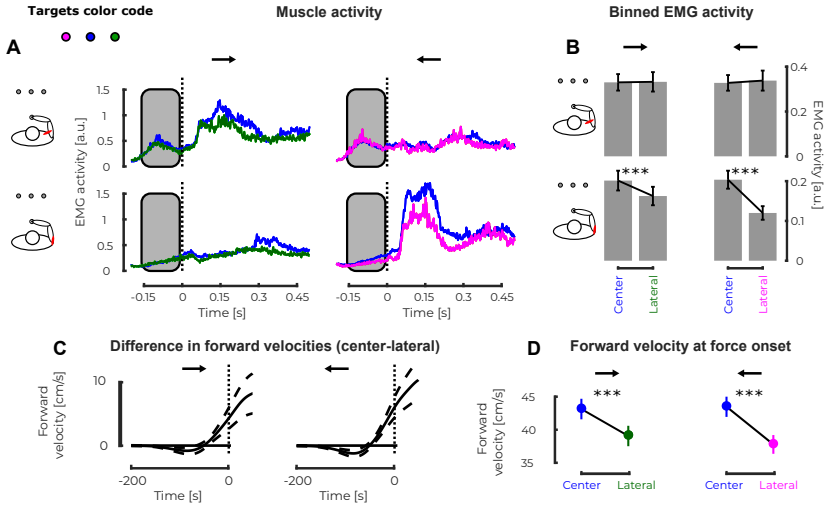

Figure IV.5: Experiment 2, EMG activity - **A** Mean EMG activity in Pectoralis Major (top) while responding to rightward (first column) and leftward perturbations (second column) and in Posterior Deltoid (bottom) while responding to rightward (first column) and leftward (second column) perturbations in the second experiment. The magenta, blue and green traces represent the mean EMG activity measured when participants reached the left, center or right target respectively. **B** Binned EMG activity before force onset in Pectoralis Major (top) and Posterior Deltoid (bottom) for the leftward and rightward perturbation loads, for the trials that reached the central (left bin) and lateral (right bin) targets. **C** Group mean and SEM of the differences in forward velocities across participants between the center and lateral trials for trials with rightward (left) and leftward (right) perturbations. **D** Comparison of the forward velocity at force onset for the trials that reached the central (blue) and lateral (green or magenta) targets with a rightward or leftward perturbation load. $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.005$ (***)

force directions (left: linear mixed model $\beta_1 = -0.019$, $t = -1.76$, $p = 0.0782$, $r^2 = 0.622$ and right: linear mixed model $\beta_1 = 0.0054$, $t = 0.89$, $p = 0.3758$, $r^2 = 0.6431$). However, we observed an increase in the EMG activity of PD prior to perturbation onset for the trials that reached the center target compared to the ones that reached the lateral targets for both force directions (Figure IV.5B bottom, left: linear mixed model $\beta_1 = 0.022$, $t = 3.78$, $p < 0.005$, $r^2 = 0.6804$ and right: $\beta_1 = -0.051$, $t = -4.804$, $p < 0.005$, $r^2 = 0.6019$). This increase in EMG activity for the trials that ended at the central target was correlated with larger forward velocities at force onset. We reported in Figure IV.5C the differences in forward velocities between the center and the lateral trials for both perturbation directions. In presence of a leftward perturbation (Figure IV.5C and D, left panels), we observed

a larger forward velocity at force onset for trials that end up at the center target compared to those that reached the lateral target (linear mixed model: $\beta_1 = -0.013, t = -3.347, p < 0.005, r^2 = 0.5439$). The same holds for trials with rightward perturbations (Figure IV.5C and D, right panels - linear mixed model: $\beta_1 = 0.040, t = 9.476, p < 0.005$, $r^2 = 0.5743$). Similar observations were reported in the *different values* condition. Indeed, we observed an increase in EMG activities in both muscles for the trials that ended up at the center target compared to those that reached the lateral target (PM linear mixed model: β_1=-0.051, t=-4.81, p<0.005, r^2=0.60 and PD linear mixed model β_1=0.022, t=-3.78, p<0.005, r^2=0.68). Moreover, some tendencies were observed in the forward speed for trials with rightward (linear mixed modelβ_1=-0.011, t=-2.019, p=0.0436, r^2=0.52) and leftward (linear mixed modelβ_1=0.012, t=1.95, p=0.0505, r^2=0.57). These results collected in the *different values* condition have to be analysed with caution because of the low number of trials that ended up at one of the two lateral targets (7.5% in the *different values* and 23.5% in the *same values* condition).

We also tested whether the reward condition (ie. *same values* and *different values*) modified movement vigor by comparing the forward velocities and muscle activities at force onset between both reward conditions. We did not observe any difference in forward velocities between both reward conditions at perturbation onset as reported by mixed effect models (t=0.60, p=0.54, r^2=0.03). Similarly, we did not observe any differences in EMG activities averaged during the 50ms preceding perturbation onset as revealed by mixed model analyses (PM: t=-0.3962, p=0.69, r^2=0.28 and PD: t=0.07, p=0.93, r^2=0.06). This absence of correlation between the reward condition and movement vigor was interesting as it confirmed that we did not introduce any experimental modulation of vigor in our paradigm. The differences in switching frequencies observed between the *same* and *different values* conditions are therefore attributable to the reward distribution and to vigor variability within both reward conditions.

This second experiment showed that humans take reward into account to respond to perturbations and potentially change target goal during movement. More specifically, participants will tend to reduce their frequency of reaching toward targets that have a lower reward. We also showed that some parameters measured before the perturbation onset are correlated with the target that is reached at the end of the trial.

3.3 Effect of the pre-activation of muscle on the motor decision

An outstanding question is when was the decision made to switch target. Did participants decide to changer after the perturbation, or did they plan to change prior to movement? On the one hand, in this experiment as in previous reports [Nashed et al., 2014], changes in goal target depend on the occurrence and magnitude of the force so it is at least partially determined by sensory information collected during movement. On the other hand, the observation that the switch also depended on the baseline activity suggests that there could be an influence of the state of the limb from the beginning of the movement on the decision. We wanted to investigate this possibility in Experiment 3. This experiment was specifically designed to investigate whether the pre-activation of PD prior to movement onset could bias the frequency of target switches. Participants had to reach any of the three targets located at the same position as in Experiment 2. All targets had the same reward in this experiment. During movement, mechanical perturbation loads could push participant's hand orthogonally to the main reaching decision. During half of the trials, a leftward background load was applied to participant's hand throughout movement evoking a background activation to counter the background load (see section 2.5). We assessed the effect of pre-activation of PD by investigating the reach proportions of the lateral targets with respect to force intensity and background condition.

The application of a leftward background force induced an increase in both PM and PD baseline EMG activity (Figure IV.6C). We found a significant effect of the background load in both muscles (main effect of the linear mixed model on both muscles: $\beta_1 = -0.11$, $t = -6.73$, $p < 0.005$, $r^2 = 0.9139$) as represented in Figure IV.6D. Moreover, we also observed an interaction effect between the background load and the muscle: baseline activity in PD increased more than PM activity ($\beta_{12} = 0.20$, $t = 19.067$, $p < 0.005$, $r^2 = 0.9139$).

We found that the leftward background load modified the reach proportion to the left target for all kind of online mechanical loads. Figure IV.6A represents the reaching proportions to the left and right targets (respectively left and right column) with respect to the intensity of the perturbation load for the trial with (bottom) or without (top) back-

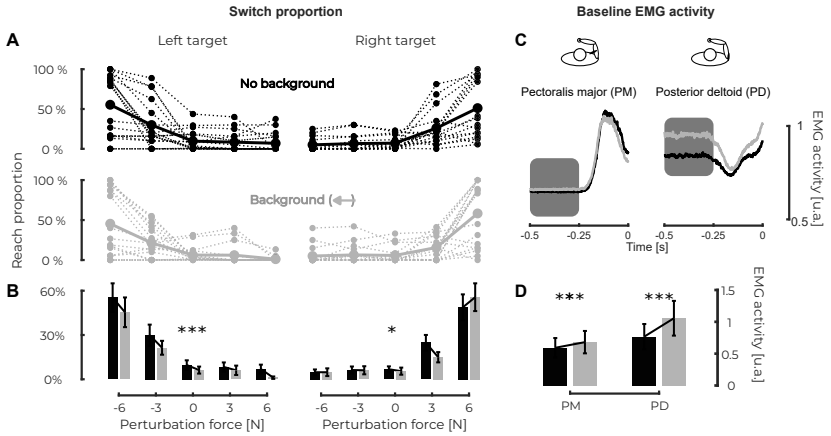

Figure IV.6: Experiment 3 – **A** Group mean (full line) and individual means (dashed lines) of the reach proportion to the left and right targets for the conditions without (top row, black) and with (bottom row, gray) the leftward background load as a function of perturbation load. **B** Comparison of the reach proportion of the left and right targets (left and right columns respectively) with (gray boxes) and without (black boxes) the leftward background force. **C** Group mean of the EMG activity in Pectoralis Major (left) and Posterior Deltoid (right) prior to movement onset for trials without (black) and with (gray) a background load. **D** Comparison of the binned EMG activity between 500 and 300ms (corresponds to the gray box in panel **C**) before force onset in Pectoralis Major and Posterior Deltoid for the conditions with and without background. $p < 0.05$ (*), $p < 0.01$, $p < 0.005$ (***)

ground load. In order to show the effect of the background load on the reach proportion of the lateral targets, we fitted a multilinear logistic regression (see section 2) that inferred the effect of perturbation and background load on the reached target. This multilinear logistic regression revealed a significant effect of the background and perturbations loads on both the left and right targets reaching proportion. Concerning the perturbation load, we observed an increase of the reach proportion of the left target with increasing leftward force ($\beta_1 = -0.9223$, $t = -22.16$, $p < 0.005$) and the mirror effect for the right target ($\beta_1 = 1.0204$, $t = 23.86$,$p < 0.005$). The background load also had a significant effect on the reach proportion for these two targets. The reach proportion to the left target decreased when the background load was applied ($\beta_1 = -0.4611$, $t = -6.1759$ and $p < 0.005$, Figure IV.6B left panel). Intuitively an increase in force towards a target should bias the choice for that target but it was not the case. A slight decrease in reach proportion for the right target was also revealed by this regression ($\beta_1 = -0.1544$, $t = -1.9972$, $p = 0.0458$, Figure IV.6B right panel).

The intensity of this effect on the two lateral targets was compared using bootstrap resampling: this effect was larger for the left than for the right target. We generated 1000 bootstrap datasets from the original dataset used to fit the multilinear logistic regression and fitted the multilinear regression on each of these bootstrap datasets (generating that way estimates of β_1 for each resampled dataset). We extracted bootstrap estimates of the main effect of background on the target reached for both lateral targets and computed the difference between the left and right estimates. The mean value of this difference was 0.280 and the 95% confidence interval obtained from bootstrap resampling was $[0.072, 0.503]$ which therefore indicates a non-zero difference. This result suggests a directional bias in the effect of the background load on the switching strategies: the application of a leftward background load hindered switches to the left target more than those to the right one.

Post-hoc analyses performed on the individual reach proportion to lateral targets confirmed this asymmetry between left and right target (see Figure IV.6B). We observed a significant decrease of the individual reach proportion to the left target induced by the background load across participants and force levels (one tailed Wilcoxon signed-rank test : $z = 2.83$, ranksum $= 999.5$, $p < 0.005$, $d = 0.21$). No similar effect was observed for the right target (Wilcoxon signed-rank test: $z = 1.23$, ranksum $= 1015$, $p = 0.2154$, $d = 0.0332$).

An interesting question is whether this background force also modulated forward velocity. We address this question by using a linear mixed model to compare forward speed at force onset in the conditions with and without background load. No modulation of movement speed between these conditions was observed (linear mixed model $\beta_1 = -0.009 \pm 0.007$, $t = -1.2528$, $p = 0.210$, $r^2 = 0.44$). This result is important as it discards the eventuality that the modulation of flexibility to switch to a new target goal was induced by movement velocity. Similarly, the reaction time was not modulated by the presence of the background force (linear mixes model, $p < 0.05$).

The results of this last experiment showed that the tendency to switch observed in Experiment 2 depended on the biomechanical state of the limb. Surprisingly, the application of a background load in a direction reduces the tendency to switch in this direction in a larger amount that the tendency to switch in the opposite direction.

4 Discussion

We conducted a series of experiments to investigate the impact of reward on feedback control strategies and rapid motor decisions by probing the impact of explicit target reward. In Experiment 1, we demonstrated that target reward did not only increase movement vigor as reported in previous studies [Summerside et al., 2018, Yoon et al., 2018], but it also increased feedback gains and muscle activity. We observed that perturbation-related lateral hand deviations were smaller when participants reached toward a target associated with higher reward. Moreover, we also observed an increase in the baseline EMG activity as well as an increase in the EMG response to perturbation loads with increasing reward. Altogether, these results suggest that the feedback gains used to perform movements scale with the value of the reward. In the second experiment, we reported that the reward distribution across competing options has an influence on rapid motor decisions: participants were less prone to switch to a nearby target if it was associated with lower reward. In addition, an increase in feedback gains was detrimental to the ability to switch target during movement as we observed in Experiments 2 and 3 that participants were also less likely to switch target during movement when the muscle activity was higher. The modulation in muscle activity introduced experimentally in Experiment 3 induced a directional bias in the ability to switch target online, demonstrating a causal influence of muscle activity.

The increase in movement vigor and feedback gains associated with reward that we observed in Experiment 1 was coherent with the use of a robust control policy [Bian et al., 2020, Crevecoeur et al., 2019]. A robust controller consists in an alternative to stochastic optimal control [Todorov and Jordan, 2002] that has the property to consider unmodeled disturbances [Basar and P, 1991] which results in better responses to mechanical perturbations during movement. Reward is known to invigorate movements as revealed in saccadic eye movements where faster movements were observed toward higher monetary rewards [Manohar et al., 2015, Manohar et al., 2017] or toward targets associated with higher implicit rewards [Xu-Wilson et al., 2009]. Similar observations were made for upper limb reaching movements that exhibited higher peak velocities toward more rewarding targets [Esteves et al., 2016, Sackaloo et al., 2014, Summerside et al., 2018, Yoon et al., 2018]. This was taken as evidence for reward-dependent

selection of movement time [Haith et al., 2012, Shadmehr et al., 2010b]. It has recently been demonstrated that this increase in movement vigor was associated with higher muscle activity in presence of reward [Codol et al., 2020] which could be interpreted as a mechanism used to increase internal feedback gains in order to improve reward-related endpoint accuracy [Manohar et al., 2019]. Here we postulate that another mechanism is also at play: a higher reward produced a more robust policy that revealed participants' will to render their movements less sensitive to perturbations, thereby reducing the risk to miss the goal. In this framework the reduction in movement time results from the robustness of the control that impacts movement velocity through larger goal-feedback gains.

In this framework, the modulation of the robustness of control has a clear limitation that we were able to establish empirically: a robust control strategy is meant to reject disturbances indistinguishably, thus in principle it is clear that this strategy is not compatible with a flexible change in movement goal online, which requires a reduction in feedback response to let the perturbation redirect one's hand toward the new goal.

Besides the property of the robust model to predict larger feedback gains, we measured here as in previous work that this strategy was associated with an increase in baseline co-activation [Crevecoeur et al., 2019] which potentially influences the gains of short- and long-latency responses to mechanical perturbations [Bedingham and Tatton, 1984, Marsden et al., 1976, Matthews, 1986, Pruszynski et al., 2009, Stein et al., 1995, Verrier, 1985]. Considering this, the competition between robust control and flexible online decisions in the human motor system may depend in part of the fact that the robust controller recruits the peripheral motor apparatus (i.e. muscle state and reflex gains) to increase the overall feedback gains, thereby creating a competition between peripheral mechanisms engaged in control and more central decisional processes.

Moreover, our results demonstrate that the model based on distributed consensus of decision-making [Cisek, 2012] also applies to online motor decisions. This framework posits that decisions occur through an competition between the different options by integrating the motor costs incurred to each action [Cos et al., 2011, Morel et al., 2017, Shadmehr et al., 2016] and their respective outcome

[Trommershäuser et al., 2003, Trommershäuser et al., 2008]. We documented a concomitant influence of both the reward distribution across competing options and load magnitudes which highlights that these two factors are taken into consideration during online motor decision. In addition, we add to these factors that the state of the peripheral motor system, influenced by the selected control strategy and feedback gains, had an effect on online decision-making. Our findings are in line with previous work reporting an impact of the cost of each action [Kurtzer et al., 2020, Nashed et al., 2014, Michalski et al., 2020] and their associated outcome [Cos et al., 2021, Martí-Marca et al., 2020]. These observations support that online motor decisions must result from distributed consensus between control strategies, feedback responses and rewards. Importantly, the present study investigated participants' decision to switch to alternative targets during movement, all of which leading to successful movements, there were no good or bad choices as it is the case in a go-before-you-know paradigm [Chapman et al., 2010, Gallivan et al., 2016b, Gallivan et al., 2017, Enachescu et al., 2021].

To conclude, our study highlights that multiple mechanisms underlie reward-dependent planning and control of movement. One the one hand, we suggest that there is a robust control strategy that involves peripheral circuits by means of increases in baseline activity and gain scaling of the feedback responses. This strategy associated with robust control is likely selected to reject perturbations and reduce the risk of missing the reward suggesting that there could be a cost incurred to reward. On the other hand, there exists a more flexible control policy able to switch target during movement. It is conceivable that this second policy, which requires some inhibition of muscle activity and response, is mediated by higher level inhibitory circuits and response modulation [Scott, 2016, Shadmehr and Krakauer, 2008]. Both policies take into account explicit target rewards and depend on the state of peripheral control loops.

An interesting open question is whether and how much individuals can modulate their strategy or whether the differences in strategy reflect individual traits. Indeed, individual differences have been shown in movement vigor [Reppert et al., 2018], and their possible effect on the

modulation of feedback control is an exciting open question. Such ability is potentially central to understand planning and control in complex environments.

Chapter V

Expanding the Optimal Feedback Control framework

It means constantly seeking and implementing better ways of thinking and acting across old and new corners of the system

Robert I. Sutton

The caveat of the traditional mathematical models used to simulate upper-limb human reaching movements is that they do not assume much flexibility for the underlying control policies. These control policies are tuned to the task goal such that they can appropriately predict feedback responses to perturbations during movement. However, the flexibility that humans demonstrate in their adaptation to rapid change in task goal and to quickly change their target during movement is not grasped by these traditional models. Here we develop an extension of the Optimal Feedback Control framework that integrates flexible changes of the control policies during movement. More specifically, we implement two different extensions that address this caveat. In the first extension, we designed a dynamical alteration of the control policy that allows to simulate movements to a target whose structure change during movement. We observed that the model is able to quickly change its control policy in response to change in target structure accordingly to what we observed experimentally. In the second extension, we develop an algorithm that allows to control movements characterized by non-

convex composite cost functions as those corresponding to movements where several alternative targets are available. We observed that the results of the algorithm are coherent with experimental behaviors as we could predict the outcome of online decision processes involving target locations and reward. These two extensions of the OFC framework allow to integrate the versatility of human behavior during movement in the existing theoretical framework.

1 Introduction

Behavioral and modeling studies provide complementary and necessary insights into humans sensorimotor control [Krakauer et al., 2017, Datta et al., 2019]. On the one hand, experimental work provides a detailed description of humans behavior through the various types of data that can be collected (kinematics, EMG, neural data, ...). However, this approach only provides a description of the agent's behavior without explaining their causes. On the other hand, mathematical modeling is a powerful tool that can be used to investigate humans behavior and to propose some assumptions for their causes. Many different mathematical models can be used to investigate the same phenomena and this redundancy requires some caution in model selection [Blohm et al., 2020].

Amongst the various control models used to explain humans behavior during reaching, Optimal Feedback Control stands out with its ability to explain feedback responses to perturbations as well as unperturbed behaviors [Todorov and Jordan, 2002, Scott, 2004, Shadmehr and Krakauer, 2008, Shadmehr et al., 2010b]. This theoretical framework has been thoroughly used to investigate, model, and explain reaching movements in various contexts [Nashed et al., 2012, Nashed et al., 2014, Izawa and Shadmehr, 2008, Diedrichsen, 2007, Kasuga et al., 2021, Omrani et al., 2013, Cesonis and Franklin, 2020, Poscente et al., 2021]. Despite being a locally outstanding candidate for reaching movements, this model relies on strict hypotheses that are not always verified when considering humans movements such as the linearity of the motor plant for instance [Sabes et al., 1998, Gordon et al., 1994, Uno et al., 1989, Schaffelhofer et al., 2015].

The most widely used implementation of OFC for human reaching movements, the Linear Quadratic Regulator (LQR), computes the feedback gains offline before movement initiation. These feedback gains are obtained from the minimization of a performance index defined as a weighted sum of motor cost and motor error (capturing target structure). Therefore, the computed feedback gains are specifically tuned to the task and the target presented before movement. This framework does not allow to adjust the feedback gains during movement in response to a change in target structure or to model participants' decision to reach for an alternative target.

Here, we propose an upgrade of this theoretical framework that allows to model dynamical properties of the reaching control policies as we observed experimentally. We first develop an algorithm that permits to model online adjustments in control policy in response to change in target structure. Then, we develop a second algorithm that leverages the notion of cost-to-go function, central concept in Optimal Control, to model online motor decisions that could occur during movement. These two algorithms produce results that are coherent with our experimental findings and provide a promising ground to build future models of human reaching movements.

2 Theoretical background

The goal of this first section is to develop, in more details than in the general introduction, the theoretical background of Optimal Feedback Control as well as the new features that we developed in the present work.

2.1 Optimal Feedback Control

Let us consider the discrete-time dynamical system with imperfect measurements of the state $\mathbf{x} \in \mathbb{R}^n$ defined by :

$$\mathbf{x}[t+1] = \mathbf{A}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \xi[t], \tag{V.1}$$

$$\mathbf{y}[t] = \mathbf{H}\mathbf{x}[t] + \omega[t], \tag{V.2}$$

where $\mathbf{A} \in \mathbb{R}^{nxn}$, $\mathbf{B} \in \mathbb{R}^{nxm}$, and $\mathbf{H} \in \mathbb{R}^{pxn}$ are constant real-valued matrices. $\xi[t] \sim \mathcal{N}(0, \Sigma_{\text{motor}})$ and $\omega[t] \sim \mathcal{N}(0, \Sigma_{\text{sensory}})$ are additive motor and measurement noises, respectively. The motor command $\mathbf{u} \in \mathbb{R}^m$ is the input that the agent can use to steer the system. The agent observes the state of the system through a noisy and possibly delayed observation $\mathbf{y} \in \mathbb{R}^p$.

To be able to steer the system towards a desired state $\mathbf{x}^*$, one has to define a performance index that captures the intended behavior and the allowed energy expenditure to reach this state. The cost-function capturing that performance index is defined based on the extended state vector $\tilde{\mathbf{x}} = [\mathbf{x}, \mathbf{x}^*]^T$ and the motor command $\mathbf{u}$ as follows

$$J(\tilde{\mathbf{x}}, \mathbf{u}) = \sum_{i=1}^{N} \left\{ \tilde{\mathbf{x}}[i]^T \mathbf{Q}[i] \tilde{\mathbf{x}}[i] + \mathbf{u}[i]^T \mathbf{R}[i] \mathbf{u}[i] \right\}, \tag{V.3}$$

where $\mathbf{Q}[t] \in \mathbb{R}^{nxn} \geq 0$ and $\mathbf{R}[t] \in \mathbb{R}^{mxm} > 0 \; \forall t \in [0, N]$ respectively capture the penalty on the state and action. This cost-function is expressed for a finite time horizon of N time steps.

For many applications, including the modeling of reaching movements, only the very last $\mathbf{Q}$-matrix is set to non-zero value as one wants to penalize motor error only for the very last time step, while the $\mathbf{R}[t]$-matrices are set to the same non-zero matrix for every time step. In addition to this general cost-function, one can define the *cost-to-go function* under the optimal policy, denoted $v^\pi(\tilde{\mathbf{x}}[t], \mathbf{u}[t])$, that captures the expected cost remaining given that the actions $\mathbf{u}^\pi[j] \; \forall j \in [t, N]$[1] are followed from state $\tilde{\mathbf{x}}[t]$[2].

This *cost-to-go* function is defined by:

$$v^\pi(\tilde{\mathbf{x}}[t], \mathbf{u}[t]) = \min_{\mathbf{u}[t]} \left\{ \tilde{\mathbf{x}}[t]^T \mathbf{Q}[t] \tilde{\mathbf{x}}[t] + \mathbf{u}[t]^T \mathbf{R}[t] \mathbf{u}[t] + \mathbb{E}[v^\pi[t+1]] \right\} \tag{V.4}$$

[1] Taken under the optimal control policy $\pi*$ in this case.

[2] The reader familiar with *Reinforcement learning* will have noted how close this definition is to that of the *state value function* $V^\pi(s)$ which is central to RL formulation. These two functions are actually opposite in sign.

An important property of this *cost-to-go* function is that, under the optimal policy π^* and for the fully observable case, it can be shown to be a quadratic function of the state vector which can be formulated as the following theorem[3]

$$v^{\pi}\left(\tilde{\mathbf{x}}\left[t\right],\mathbf{u}\left[t\right]\right)=\tilde{\mathbf{x}}\left[t\right]^{T}\mathbf{S}\left[t\right]\tilde{\mathbf{x}}\left[t\right]+s\left[t\right], \tag{V.5}$$

where $\mathbf{S}\left[t\right]\in\mathbb{R}^{nxn}\geq 0$ and $s\left[t\right]$ are non-negative quantities.

During reaching movements, the state of the system is not fully observable as the agent observes the environment through vision and proprioception which are noisy and delayed versions of the state. In this case, the result developed in equation V.5 can be reformulated based on the state estimation $\hat{\tilde{\mathbf{x}}}\left[t\right]$ instead

$$v^{\pi}\left(\hat{\tilde{\mathbf{x}}}\left[t\right],\mathbf{u}\left[t\right]\right)=\hat{\tilde{\mathbf{x}}}\left[t\right]^{T}\mathbf{S}\left[t\right]\hat{\tilde{\mathbf{x}}}\left[t\right]+s\left[t\right] \tag{V.6}$$

The forward recursion performing the optimal state estimation thanks to a Kalman filter writes as follows

$$\hat{\tilde{\mathbf{x}}}\left[t\right]=\mathbf{A}\hat{\tilde{\mathbf{x}}}\left[t\right]+\mathbf{B}\mathbf{u}\left[t\right]+\mathbf{K}\left[t\right]\left(\mathbf{y}\left[t\right]-\mathbf{H}\hat{\tilde{\mathbf{x}}}\left[t\right]\right), \tag{V.7}$$

$$\mathbf{K}\left[t\right]=\mathbf{A}\boldsymbol{\Sigma}\left[t\right]\mathbf{H}\left(\mathbf{H}\boldsymbol{\Sigma}\left[t\right]\mathbf{H}^{T}+\boldsymbol{\Sigma}_{\text{sensory}}\right)^{-1}, \tag{V.8}$$

$$\boldsymbol{\Sigma}\left[\boldsymbol{t+1}\right]=\boldsymbol{\Sigma}_{\text{motor}}+\left(\mathbf{A}-\mathbf{K}\left[t\right]\mathbf{H}\right)\boldsymbol{\Sigma}\left[t\right]\mathbf{A}^{T}, \tag{V.9}$$

Under this assumption, the optimal motor commands (i.e. those that minimize equation V.3) are defined by

$$\mathbf{u}^{*}\left[t\right]=-\mathbf{L}\left[t\right]\hat{\tilde{\mathbf{x}}}\left[t\right], \tag{V.10}$$

where the feedback gains $\mathbf{L}\left[t\right]$ are defined recursively by equations V.11-V.15 and $\hat{\tilde{\mathbf{x}}}\left[t\right]$ is the optimal state estimation at time t defined by equations V.7-V.9.

$$\mathbf{L}\left[t\right]=\left(\mathbf{R}+\mathbf{B}^{T}\mathbf{S}\left[t\right]\mathbf{B}\right)^{-1}\left(\mathbf{B}^{T}\mathbf{S}\left[t\right]\mathbf{A}\right) \tag{V.11}$$

$$\mathbf{S}\left[t-1\right]=\mathbf{Q}\left[t\right]+\mathbf{A}^{T}\mathbf{S}\left[t\right]\left(\mathbf{A}-\mathbf{B}\mathbf{L}\left[t\right]\right) \tag{V.12}$$

$$s\left[t\right]=s\left[t-1\right]+tr\left(\mathbf{S}\left[t-1\right]\Sigma_{\text{motor}}\right) \tag{V.13}$$

$$\mathbf{S}\left[N\right]=\mathbf{Q}\left[N\right] \tag{V.14}$$

$$s\left[N\right]=0 \tag{V.15}$$

[3]This theorem is only valid when considering a linear system as described by equations V.1 and a quadratic cost-function V.3 with Gaussian additive noise.

2.2 Limb dynamics

In the present work, we assume that upper limb reaching movements can be modeled using a point-mass model governed by Newton's second law of motion. This linear plant dynamics writes as follows

$$m\ddot{p} = -G\dot{p} + F + F_{ext}, \tag{V.16}$$

where p is the x- or y-position, m is the mass of the hand, G is a damping factor, F is the controlled force, and F_{ext} is the external force (that can be used to model mechanical perturbations).

The controlled force is related to the motor command by the following relationship

$$\tau\dot{F} = u - F, \tag{V.17}$$

where τ is a time constant related to the muscle dynamics that acts as a low pass filter of the motor command u. The continuous linear system can be transformed and discretized using Euler method to be written in the following discrete-time form

$$\mathbf{x}\left[t+1\right] = \mathbf{A}\mathbf{x}\left[t\right] + \mathbf{B}\mathbf{u}\left[t\right] + \xi\left[t\right], \tag{V.18}$$

where $\xi\left[t\right]$ is some additive Gaussian noise characterizing motor noise.

The state vector is defined by $\mathbf{x} = \left[x, y, \dot{x}, \dot{y}, F_x, F_y, F_x^{\text{ext}}, F_y^{\text{ext}}\right]^T$ and can be extended to integrate a non-zero target state $\tilde{\mathbf{x}} = \left[\mathbf{x}, \mathbf{x}^*\right]^T$.

In order to model the sensory delays, the state vector can be extended with past iterations (see [Crevecoeur et al., 2016] for more details). The equation that characterizes the state observation can therefore be written

$$\mathbf{y}\left[t\right] = \begin{bmatrix} \mathbf{y}_{\text{proprio}}\left[t\right] \\ \mathbf{y}_{\text{visual}}\left[t\right] \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}\left[t - \delta_{\text{proprio}}\right] \\ \tilde{\mathbf{x}}\left[t - \delta_{\text{visual}}\right] \end{bmatrix} = \mathbf{H}\tilde{\mathbf{x}}\left[t\right], \tag{V.19}$$

where the observation matrix $\mathbf{H}$ can be further developed as follows,

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}^n & \dots & \mathbf{H}_{\text{proprio}} & \dots & \mathbf{0}^n \\ \mathbf{0}^n & \dots & \mathbf{0}^n & \dots & \mathbf{H}_{\text{visual}} \end{bmatrix}, \tag{V.20}$$

where $\mathbf{0}^n$ is the null matrix of appropriate dimension and $\mathbf{H}_{\text{proprio}}$ and $\mathbf{H}_{\text{visual}}$ are the observation matrices for proprioception and vision, respectively (see [Kasuga et al., 2021] for a coherent way to fill them in[4]).

We intentionally choose to use a linear plant dynamics in the present work because we do not think it will impact the observed behavior as we mainly manipulate the cost-function. However, the non-linearity of the motor plant is not always something that can be set aside as it probably has an influence on some elements of the behavior.

3 Receding horizon

This purpose of this section is to provide some modeling support to the findings of chapters II and III by modeling the behavior that were observed in the corresponding studies.

The main topic of this section is therefore the notion of *receding horizon* [Kwon et al., 1982] which characterizes a control system that repeatedly and recurrently solves the problem associated with the phenomenon it wants to model. In the context of this thesis, it means that it considers that the time horizon associated with the movement is continuously shifting in time. Here, we develop a modeling algorithm that allows to model *receding horizon* with Optimal Feedback Control and apply it to reaching movements in order to reproduce the results of chapters II and III.

3.1 Modeling algorithm

We modified the cost-function associated with the task to model movements to target of different structures. More specifically, we altered the structure of the $\mathbf{Q}[N]$ matrix to mimic change in target structure. We first modeled the experimental results presented in chapter II, namely the impact of a switch in target width during movement. For this purpose, we considered four different types of trials. Two of them did not include any change in target structure (movements toward a wide or a narrow target) and the other two included switch from a narrow to a

[4] In the present work, we were mostly interested in the cost-functions and therefore used canonical identity matrices for each sensory observation matrix.

wide target or vice versa. The feedback gains of the unperturbed trials were computed using the backward recursion described in equations V.11 to V.15 as these trials did not involve any online change in target.

To model the trials in which a switch in target width occurred during movement, we had to dynamically compute the feedback gains in order to integrate the change in target during movement. For this purpose, we designed and used the following algorithm for the dynamical computation of the feedback gains

Algorithm dynamical computation of feedback gains

Initialize $\mathbf{L}$ and $\mathbf{S}$
for $i = \mathrm{N} : 1$ **do**
  Initalize $\mathbf{L}_i$ and $\mathbf{S}_i$ for $N_i \leftarrow N - i$ time steps
  $\mathbf{S}_i\,[\mathrm{end}] \leftarrow \mathrm{QN}_i$
  Perform the backward recursion on $\mathbf{L}_i$ and $\mathbf{S}_i$
  $\mathbf{L}\,[i] \leftarrow \mathbf{L}_i[1]$
  $\mathbf{S}\,[i] \leftarrow \mathbf{S}_i[1]$
end for

This algorithm consists in a modification of the backward recursion introduced above. The matrices $\mathbf{L}\,[t]$ and $\mathbf{S}\,[t]$ are computed recursively backward. For each time step i, starting from the last one, the classical backward recursion is computed for $N - i$ time steps and only the very first feedback gain matrix $\mathbf{L}$ is kept.

Thanks to the recursivity of this computation, we can alter the cost function during the backward recursion by changing the QN_i matrices introduced at each time step. To model the switch conditions we introduced the cost matrix corresponding to the initial target for the first half of the iterations and that one corresponding to the final target for the second half of the iterations.

These cost matrices were defined such that they put penalty on the state vector $\tilde{\mathbf{x}} = [x,\, y,\, \dot{x},\, \dot{y},\, x^*,\, y^*,\, \dot{x}^*,\, \dot{y}^*]$ capturing the state of the dynamical system at any time

$$\text{QN} = \begin{bmatrix} w_1 & 0 & 0 & 0 & -w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 & 0 & -w_2 & 0 & 0 \\ 0 & 0 & w_3 & 0 & 0 & 0 & -w_3 & 0 \\ 0 & 0 & 0 & w_3 & 0 & 0 & 0 & -w_3 \\ -w_1 & 0 & 0 & 0 & w_1 & 0 & 0 & 0 \\ 0 & -w_2 & 0 & 0 & 0 & w_2 & 0 & 0 \\ 0 & 0 & -w3 & 0 & 0 & 0 & w_3 & 0 \\ 0 & 0 & 0 & -w_3 & 0 & 0 & 0 & w_3 \end{bmatrix} \tag{V.21}$$

We set $w_2 = 100$ for the *narrow* target and $w_2 = 0$ for the *wide* target. Therefore, the *narrow to wide* (*wide to narrow*) condition started the movement with $w_2 = 100$ ($w_2 = 0$) and changed it to $w_2 = 0$ ($w_2 = 100$) at the middle of the movement to mimic the switch in target goal. The results of this modeling work are presented in section 3.2.1.

In order to model the results of chapter III, we used the same algorithm but we altered the modifications of the QN_i matrices. Indeed, in the *slow* and *fast* conditions we had to introduce continuous changes in the cost-function during movement since the target width was continuously changing in these conditions. We did this by changing the value of w_2 during movement using the following time-varying functions[5]

$$w_{2,i}^{\text{slow}} = \begin{cases} 0, & \text{if } i < 15 \\ \dfrac{0.001 w_2^{\text{narrow}}}{1 + \exp\left\{-\dfrac{(i-10)}{3}\right\}}, & \text{if } i >= 15 \end{cases} \tag{V.22}$$

$$w_{2,i}^{\text{fast}} = \begin{cases} 0, & \text{if } i < 15 \\ \dfrac{0.01 w_2^{\text{narrow}}}{1 + \exp\left\{-\dfrac{(i-10)}{3}\right\}}, & \text{if } i >= 15 \end{cases} \tag{V.23}$$

The results of this modeling work are presented in section 3.2.2.

[5]These functions were obtained by trial-error and might not be the best ones to mimic the experimental behaviors.

3.2 Results

3.2.1 Instantaneous modification of target width

Using the methodology described in section 3.1, we were able to reproduce the experimental results collected in chapter II. We modeled reaching movements to a target that was initially a square (red in figure V.1A) or a rectangle (blue in figure V.1A) whose width could suddenly change during movement (dashed lines in figure V.1A). We observed that these changes in target width impact the observed behavior in presence of mechanical perturbations (to the right in figure V.1). Similarly to what we observed experimentally, the switch from a *narrow* to a *wide* target induced an increase in the lateral hand deviation compared to the *narrow* condition (red dashed vs full lines in figure V.1C). Inversely, the switch from *wide* to *narrow* target induced the mirror effect on the behavior (blue dashed vs full lines in figure V.1C). In both cases, we did not observe any modulation of the behavior along the main reaching direction (figure V.1B) which confirmed our experimental findings.

We also investigated the observed modulation of the motor command which is reminiscent of the modulation of the EMG activity of the Pectoralis Major. Indeed, both the x-motor command and PM strive to counteract rightward perturbation. We observed in our simulations an increase (decrease) in the *x-motor command* when the target turned from a *wide* to a *narrow* (*narrow* to *wide*) (see figure V.1E) similarly to the PM activity we reported in Chapter II.

3.2.2 Continuous modification of target width

The algorithm we developed above allows us to reproduce the results we observed in Chapter III: the impact of continuously changing targets on the reaching behavior. Indeed, we could demonstrate that the reaching behavior depends on the rate of change in target width (see green and blue traces in figure V.2A for the *slow* and *fast* conditions). Coherently with our experimental results, our model predicted that the lateral hand deviation (figure V.2C) in the *slow* and *fast* conditions are between the *switch* and the *no change* conditions. The inspection of the predicted motor command (figure V.2E) confirms the modulation of the EMG activity in the muscles stretched by the mechanical per-

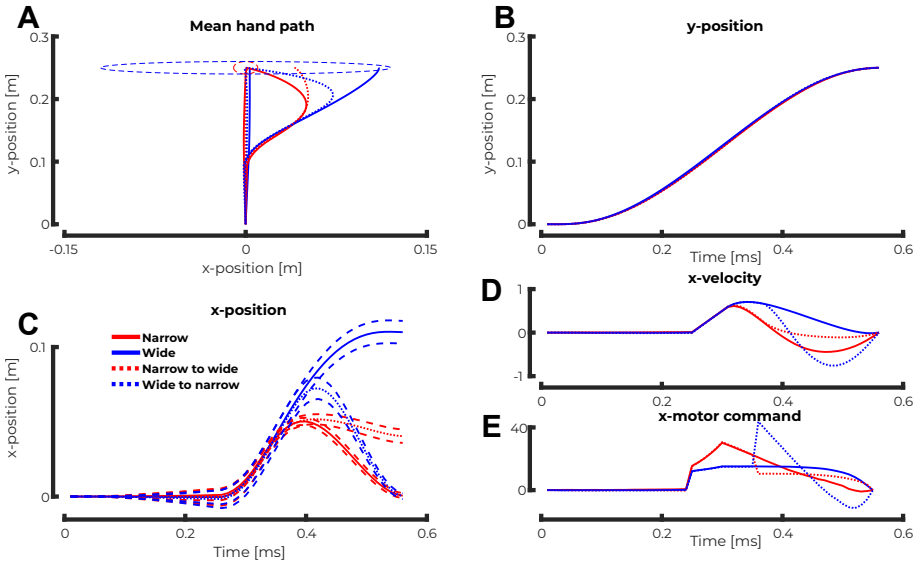

Figure V.1: A Group mean of the hand path trajectory to the narrow (red) and wide (blue) goal target in presence or absence of a rightward mechanical load. The dashed lines represent the trials in which a visual perturbation changed the structure of the initial target from narrow to wide (red dashed line) or wide to narrow (blue dashed line). The dashed ellipses at the top of the panel represent the different targets and their respective width. **B** Group mean of the y-position as a function of time for the different conditions. **C** Group mean and standard deviation of the x-position as a function of time for the different target conditions. The full lines correspond to the trials without changes in target (blue, wide and red, narrow) while the dashed lines represent those we the structure of the target was altered (blue, wide to narrow and red narrow to wide). **D** Group mean of the x-velocity for the different target conditions. **E.** Group mean of the x-motor command for the different target conditions.

turbation. Indeed, we observed a difference between the *slow* and *fast conditions* indicating that the different rate of changes in target width were responded differently by the controller.

3.3 Intermediate discussion

In this section, we demonstrated that a receding horizon controller can reproduce the experimental results of chapters II and III which the finite horizon formulation of the Optimal Feedback Control could not do. This confirms the existence of mechanism able to adjust the control policy during movement to rapidly adjust the behavior to change in

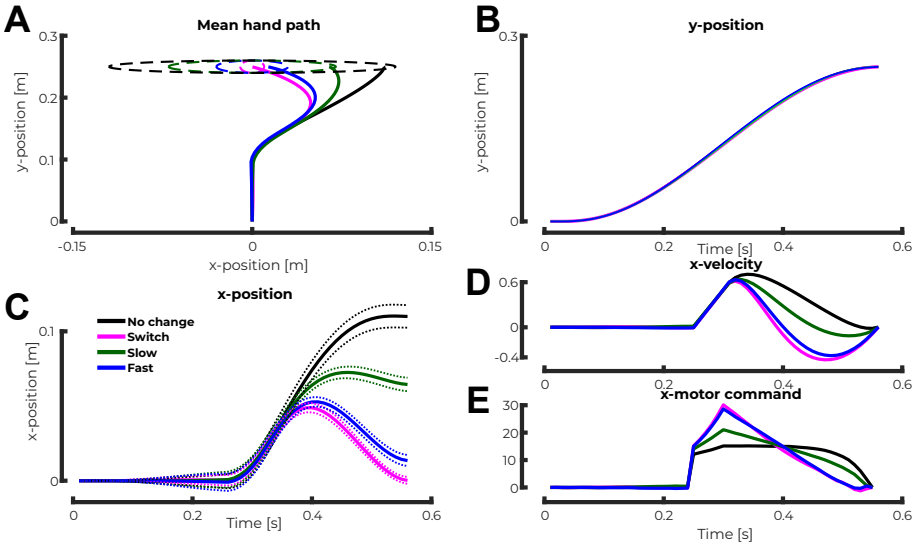

Figure V.2: A Group mean of the hand path trajectory to the different targets in presence of a rightward mechanical perturbation. The black trace corresponds to the *no change* conditions, the magenta one to the *switch* condition, the blue one to the *fast* condition and the green one to the *slow* condition. The ellipses at the top of the panel represent the different targets and their respective width at the end of movement. **B** Group mean of the y-position as a function of time for the different target conditions. **C** Group mean and standard deviation of the x-position a function of time for the different target conditions. **D** Group mean of the x-velocity for the different target conditions. **E** Group mean of the x-motor command for the different target conditions.

task constraints. Moreover, this mechanism is sensitive to the specific change in task demands as demonstrated by the modulation of the behavior in the *slow* and *fast* conditions.

Even though, using this receding horizon framework, we could expand the OFC framework by adding some flexibility in the range of behavior that can be investigated with the model, it remains a gross approximation of what the central nervous system. Indeed, it relies on extremely strong assumptions such as the linear motor plant that is controlled[6], the quadratic cost-function and the full knowledge of the plant dynamics.

[6]This is the reason why only the motor command could be investigated here. Some methods exist to model non-linear plant dynamics by linearizing them around nominal trajectories. We tried to avoid these implementation as they rely on the additional assumption of trajectory planning.

4 Online motor decisions

In chapter IV, we extended the distributed consensus model developed by Cisek [Cisek, 2012] to online motor decisions that occur during movements. An interesting question is whether such decisions can be modeled using the existing theoretical framework of Optimal Feedback Control (OFC). The goal of this section is to explore this possibility through the development of a control algorithm designed to navigate composite cost-functions. The results of the modeling work will be compared to the experimental results and a small perspective of their meaning will wrap up this section.

4.1 Modeling algorithm

The goal of this section is to detail the implementation we used to model online motor decisions between different targets using the OFC framework.

To model reaching movements to different targets, we computed the backward recursion described by equations V.11 to V.15 for each of them. Since both the feedback matrices $\mathbf{L}[t]$ and the $\mathbf{S}[t]$ matrices are independent on the target location, we obtained exactly the same values for each target. The difference in the target locations matters when we consider the optimal motor command at any time step defined by

$$\mathbf{u}^*[t] = -\mathbf{L}[t]\,\hat{\mathbf{x}}[t], \tag{V.24}$$

where $\hat{\mathbf{x}}[t]$ is the state estimate at time t. This state estimate mirrors the latent state of the system which contained target location. Since the target location is contained within that state estimate[7], one has to define the state and its estimate such that each and every alternative target is contained in the state vector. For this purpose, we extended the state vector as follows

$$\tilde{\mathbf{x}} = \begin{bmatrix}\mathbf{x} & \mathbf{x}^{*1} & \dots & \mathbf{x}^{*j}\end{bmatrix}^T \tag{V.25}$$

where $\mathbf{x}^{*i}$ corresponds to the i^{th} target. The $\mathbf{L}[t]$ and $\mathbf{S}[t]$ matrices for each target will have the same non-zeros entries that will be shifted depending on the target they refer to. For instance, the feedback matrices $\mathbf{L}[k]$ can be written, in block notation, as follows for three different

[7] Because the state estimate $\hat{\mathbf{x}}$ estimates the latent state $\tilde{\mathbf{x}} = [\mathbf{x}, \mathbf{x}^*]^T$

targets

$$\mathbf{L}^1 = \begin{bmatrix} \mathbf{L} & -\mathbf{L} & \mathbf{0}^n & \mathbf{0}^n \end{bmatrix} \tag{V.26}$$
$$\mathbf{L}^2 = \begin{bmatrix} \mathbf{L} & \mathbf{0}^n & -\mathbf{L} & \mathbf{0}^n \end{bmatrix} \tag{V.27}$$
$$\mathbf{L}^3 = \begin{bmatrix} \mathbf{L} & \mathbf{0}^n & \mathbf{0}^n & -\mathbf{L} \end{bmatrix} \tag{V.28}$$

where $\mathbf{0}^n$ is the null matrix of appropriate dimensions.

The feedback gain matrix $\mathbf{L}^i$ to be applied at each time step as well as the associated goal target to reach for are determined based on the *cost-to-go* function. Whichever target has the lowest associated *cost-to-go* function is selected as the target to reach for. These different *cost-to-go* functions were computed as follows

$$v^1\left(\tilde{\mathbf{x}}\left[k\right], \mathbf{u}\left[k\right]\right) = \tilde{\mathbf{x}}\left[k\right]^T \mathbf{S}^1 \tilde{\mathbf{x}}\left[k\right] + s^1\left[k\right] \tag{V.29}$$
$$v^2\left(\tilde{\mathbf{x}}\left[k\right], \mathbf{u}\left[k\right]\right) = \tilde{\mathbf{x}}\left[k\right]^T \mathbf{S}^2 \tilde{\mathbf{x}}\left[k\right] + s^2\left[k\right] \tag{V.30}$$
$$v^3\left(\tilde{\mathbf{x}}\left[k\right], \mathbf{u}\left[k\right]\right) = \tilde{\mathbf{x}}\left[k\right]^T \mathbf{S}^3 \tilde{\mathbf{x}}\left[k\right] + s^3\left[k\right] \tag{V.31}$$
$$\tag{V.32}$$

Together, the algorithm implemented in the modeling work can be written as follows in pseudocode

Algorithm online motor decisions

Compute the feedback gains for each target $\mathbf{L}^1$, $\mathbf{L}^2$ and $\mathbf{L}^3$
for $i = 1 : \mathrm{N}$ **do**
  $\mathbf{L}\left[i\right] \leftarrow \min\left\{v^1\left(\tilde{\mathbf{x}}\left[i\right], \mathbf{u}\left[i\right]\right), v^2\left(\tilde{\mathbf{x}}\left[i\right], \mathbf{u}\left[i\right]\right), v^3\left(\tilde{\mathbf{x}}\left[i\right], \mathbf{u}\left[i\right]\right)\right\}$
  $\tilde{\mathbf{x}}\left[i+1\right] = \mathbf{A}\tilde{\mathbf{x}}\left[i+1\right] - \mathbf{B}\mathbf{L}\left[i\right]\hat{\tilde{\mathbf{x}}}\left[i\right] + \text{noise}\left[i\right]$
end for

Concretely, this algorithm consists in defining a composite cost-function for the optimal control problem. This composite cost-function is defined as the minimum the individual functions associated with each target which are quadratic functions. Taking the minimum of these convex functions implies that the overall cost function is not convex anymore which violates the hypotheses underlying Optimal Feedback Control. The algorithm we implemented here provides the agent with a rule of thumb to navigate this concave composite cost-function. This

algorithm could be theoretically used for other applications that contains such composite cost-functions. In order to integrate different target rewards in our simulations, we modulated the values of the s^i parameters. Putting larger values for more rewarding target results in biasing the decision in that direction.

4.2 Results

Using the methodology described in section 4.1, we modeled a reaching task similar to the Experiment 2 of chapter IV. Briefly, participants had to reach to any of the three targets presented in front of them. The three targets were aligned with the x-axis and the central target was therefore closer to the home target that the other two (see figure V.3). We simulated three different experiments to investigate the impact of the relative rewards associated with the different targets on the agent's behavior. In the first experiment, all targets had the same reward (*same rewards* condition) whereas in the other two experiments the reward of the central target was larger than the other two with a larger value in the third experiment (*central huge* condition) than in the second one (*central large* condition).

In order to investigate the agent's ability to switch target we investigated the frequency at which each lateral target was reached as a function of the intensity of the perturbation load for different reward conditions. To mirror the behaviors observed experimentally, we selected as baseline condition the one where all the targets had similar reward and compared the reach frequency to the lateral targets when these had smaller reward than the central target. Figure V.3 represents an example of the obtained individual behaviors.

Figure V.4 represents the modulation of the reach frequencies to the two lateral targets as a function of the intensity of the perturbation load and the reward condition.

Interestingly, our model produce similar results that those observed experimentally. Indeed, we observed in the *same rewards* condition that the introduction of a perturbation load in a direction increased the reach proportion of the corresponding target. Moreover, the increase in reward of the central target tends to decrease the frequency of reach to the lateral targets which is also coherent with what we observed experimentally.

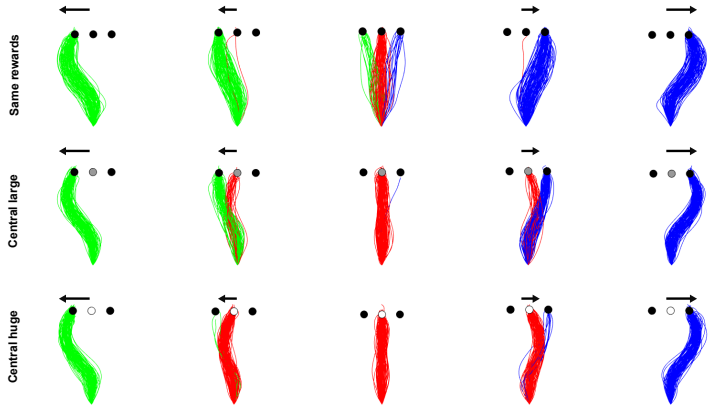

Figure V.3: Modeling results for the online decision making algorithm when reaching to three potential targets with the same rewards (top row), a larger reward for the central target (middle row) and an even larger reward for the central target (bottom row). The different columns represent the different force intensities and directions (from left to right -10N, -5N, 0N, 5N and 10N). The colors traces represent movement to the left (green), central (red) and right (blue) targets.

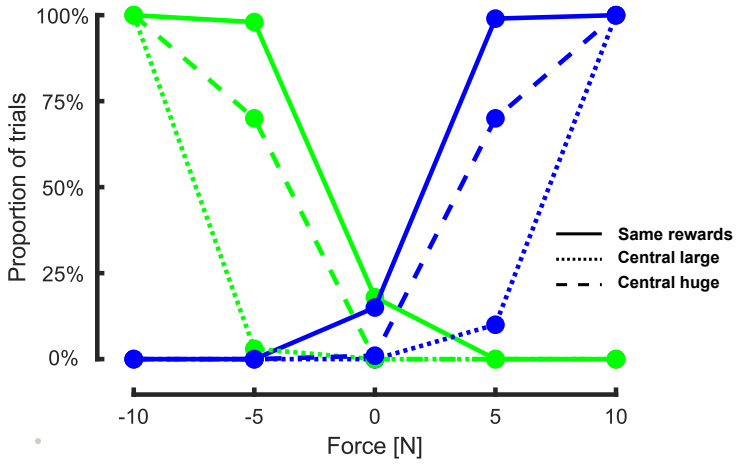

Figure V.4: Reach frequencies to the two lateral targets (green and blue for left and right, respectively) as a function of the intensity of the intensity of the perturbation force for the *same rewards* (full line), *central large* (dashed line) and *central huge* (dotted line) conditions.

In order to test the applicability of this modeling approach to other experimental data that did not imply reward, we adapted our implementation to reproduce some of the work investigated by Michalski and colleagues [Michalski et al., 2020]. We modeled reaching movements

in which the agent was initially reaching to a target located in front of them (0° in figure V.5). As they were moving, an alternative target was presented at one of the different possible angular position. We investigated the agent's probability to switch to this alternative target during movement by representing their behavior in figure V.5.

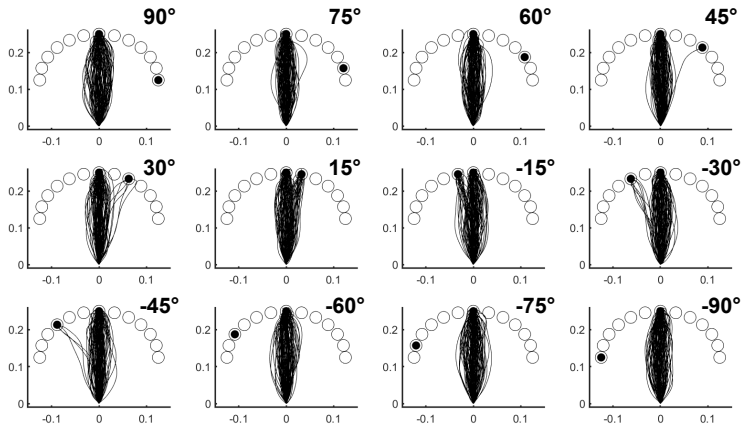

Figure V.5: Modeling results for the online decision making algorithm for an agent reaching to a target located in front of them (0°C) and that were presented an alternative target as they were at the middle of the movement. The alternative target is presented at different angular positions indicated in each panel. The black traces represent one realisation of the agent's behavior.

We observed that the probability of switch in this condition is higher when the angular differences between the initial target and its alternative is low coherently with what was observed in the Michalski's paper [Michalski et al., 2020].

4.3 Intermediate discussion

These modeling results demonstrate that online motor decisions are sophisticated processes that can be modeled using the Optimal Feedback Control framework by leveraging the *cost-to-go* function. We confirmed with this modeling work that these online motor decisions can be studied in the distributed consensus framework [Cisek, 2012]. Online motor decisions therefore not only compare the different actions but also take into account their outcome (as demonstrated by the impact of reward) and their associated biomechanical motor cost (as demonstrated by the influence of the locations of the different targets.)

By expanding the model formulation developped by Nashed and colleagues [Nashed et al., 2014] we demonstrated the ability of OFC to not only model reaching movements but also to model online motor decisions. In the present section, we demonstrated that high-level goals such as reward can be integrated within the theoretical framework of OFC to reproduce bias in motor decisions. Other experimental results implying the combination of motor control and decision making can be obtained using similar implementation.

Even though our implementation provides promising insight in the process underlying online motor decisions, some discrepancy with the experimental behaviors have to be kept in mind. Indeed, our model assumes that the agent is an optimal decision maker which means that, at any time, it can take the very best decision which is definitely not the case experimentally. Moreover, our model did not factor the voluntary intent that could influence participants' behavior as they see a new target appearing (e.g. more exploratory behavior). Other factors that are not factored in this model are the potential cognitive cost associated with the switch to an alternative target or the time that this process would take (here we assume that this process is instantaneous). Nevertheless, the behaviors we could model with this implementation are coherent with those observed experimentally which make plausible the implementation of such a mechanism based on the *cost-to-go* function underlying online motor decisions in humans.

5 Discussion and limitations

In this chapter, we developed and expanded the Optimal Feedback Control framework to model and explain behavioral results that could not be explained by its standard formulation used in the sensorimotor control field [Todorov and Jordan, 2002, Scott, 2004]. Indeed, this standard formulation assumes that the cost-function associated with the movement is defined beforehand and cannot be changed during movement. Moreover, this model cannot explain and simulate motor decisions that occur during movements such as participants' will to reach to an alternative target. Here we develop two different modeling algorithm to tackle these problems and demonstrate that they are appropriate for expanding the OFC framework to these phenomena.

To model participants' ability to adjust their behavior to online change in task demands, we developed a recursive recomputation of the feedback gains based on the model predictive control (see [Lee, 2011] for review). This implementation permits to reproduce the experimental behaviors of chapters II and III namely the online modification of control policies in response to changes in target structure. Besides the specific cases developed in the present thesis, this algorithm can be used to model any switches between different movement that can individually be modeled by OFC. For instance, it can also be used to model appearance or disappearance of obstacles during movement provided that they can be modeled thanks to a via-point as in [Nashed et al., 2012] or to model any online modification in target structure during movement.

Online motor decisions between alternative actions were modeled by leveraging the notion of cost-to-go function. This function defining the expected remaining cost was defined for each target and biased depending on the reward associated with the target. By assuming that the agent will try to reach for whichever target is associated with the lowest cost-to-go function, we were able to reproduce the experimental results of chapter IV. Indeed, we demonstrated that the reaching frequency to the lateral targets decreased when their associated reward decreased relative to the central target coherently with what we observed experimentally. Moreover, we demonstrated that this modeling approach can be used to model other motor decisions such as those that take place when an alternative target appears during movement such as in a recent study by Michalski and colleagues [Michalski et al., 2020].

An interesting question is to what extent these modeling algorithms generalize to other class of reaching movements. More specifically, we may wonder whether the algorithm we used for online motor decision can be used to model target redundancy. Indeed, in section 4.1 we developed an algorithm allowing to model movements where two or more alternative targets result in correct movements, which could be extended to any redundant target. For this purpose, the redundant target can be considered as an large set of punctual targets for which the algorithm described in section 4.1 can be applied. The main difference with the results we developed in the present thesis is that only the extended state will be modified, the feedback gains will be the same for each punctual target. The difference between the different cost-to-go functions will be mostly governed by the relative

distance between the hand and the different punctual targets. On paper, this method could be used to model reaching movements to a redundant target such as the wide targets we often used in our experiments. However, the computational cost associated with the selection, at each time step, of the best option increases a lot with the number of alternative targets which is not really coherent with humans ability to rapidly integrate target redundancy in their feedback responses [Lowrey et al., 2017, Nashed et al., 2014, De Comite et al., 2021].

The implementation of the algorithm we used to model online motor decisions is reminiscent of the field of reinforcement learning (RL) where an agent learns to control a system in an unknown environment through trial and error [Sutton and Barto, 2018]. Indeed, in reinforcement learning, the goal of the agent is to end up with an approximation of the value function, opposite in sign to the cost-to-go function of the OFC framework, that could be obtained either as a lookup table or a non-linear approximation by a neural network. The main difference between OFC and RL derivations of the optimal control policy is that in the former the cost-to-go function is approximated by applying a backward recursion starting from the final state while in the latter it is approximated through trial and error. Interestingly, the assumptions underlying the RL derivation are less restrictive as they do not assume a linear dynamic of the system and a quadratic cost function. This property allows that framework to theoretically control any system [Lillicrap et al., 2019, Levine et al., 2018, Mnih et al., 2015, Schulman et al., 2016]. Further modeling studies should investigate whether the alternative of reinforcement learning is a good candidate to derive goal-directed control policies similar to that observed experimentally.

To sum up, the modeling algorithms developed in this chapter allow to expand the theoretical framework used to model reaching movements by adding dynamical components to the control policies. These modeling results reinforce the strength of the experimental results of the previous chapters by providing theoretical ground for their interpretation. The modifications that we brought to the OFC framework consist in dynamical alteration of the control policy during movements resulting in an increase of the flexibility of the behavior characterised by the ability to adjust to online change in task demands or to change target during movements. Despite being able to model dynamical factors of

the reaching control policies, these additions to the OFC framework do not resolve all the caveats mentioned in section 2.4. For instance it does not help to discriminate between finite and infinite time horizon and does not propose a proper way to model the impact of target redundancy and reward on the behavior. The path towards a complete and biologically coherent control model of humans movement still has a promising future.

Chapter VI

Conclusion and perspectives

It is good to have an end to journey toward; but it is the journey that matters, in the end

Ursula K. Le Guin

This thesis is in the continuity of decades of research on the control policies underlying reaching movements in humans and non-humans primates. More specifically, the main theme of this work was to investigate the complex entanglement between movement planning and movement execution by exploring the impact of the parameters integrated during movement planning on movement execution. The following paragraphs summarize the different findings presented in this thesis and detail their contribution to the field of sensorimotor control.

1 Main contributions

In Chapter II, we demonstrate that the control policies used by humans to perform reaching movements are not immutable as the theoretical separation of movement planning and movement execution suggests. We report online changes in control policy elicited by changes in target structure or environmental context. More specifically, we observed that participants were able to adjust their behavior and feedback responses to mechanical perturbations following changes in task demands during movement. These changes in control policy were coherent with

the changes in task demands as we reported for instance an increase in feedback gains when the target became narrower and the inverse effect when it became wider. We further demonstrate that these adjustments in control policy were time-locked to the onset of the visual perturbation and we report an additional delay for the adjustments associated with changes in obstacles suggesting a different processing of the two conditions. To sum up, we report the existence of a feedback mechanism in the selection of human control policy that could accommodate change in task demands during movement, even in the absence of mechanical perturbations.

In Chapter III, we further investigate that feedback mechanism and demonstrate that it is able to integrate dynamical changes in task demands. This additional results is crucial as it allows to discard the eventuality that humans were computing the different control policies beforehand and switched to the appropriate one during movement. Moreover, we demonstrate that this feedback mechanism is finely tuned to the rate of change in task demands which confirm the hypothesis whereby it is mediated by a dynamical feedback loop conveying task-related information to the policy selection process. Interestingly, these online adjustments in control policy were improved across trials. Indeed, as participants get more familiar with the task, they demonstrated more appropriate adjustments in control suggesting that this feedback process could also, somehow be optimized.

In Chapter IV, we document the impact of target reward on the feedback reaching control policy demonstrating that reward could modulate their robustness. We hypothesize that higher target reward consists in more appealing targets that participants will be less willing to miss and therefore select a control policy that is better at rejecting unexpected disturbances. Elaborating on that, we suggest that the increase in robustness, characterized by an increase in feedback gains, would be detrimental of the ability to be flexible during movement. We report a correlation between higher feedback gains, attributed to more robust control policy and a reduction of flexibility during movement. This correlation was strengthened by experimentally upregulating the feedback gains and demonstrating a causal link between high feedback gains and a reduced flexibility during movement. We therefore unveil a competition between the robustness of the control policy and its flexibility.

In Chapter V, we implement two modeling algorithms to extend the existing models and simulate the experimental results of chapters II, III, and IV. The first algorithm manipulates the computation of feedback gains such that the controller can adjust to online modification in target structure. This algorithm reproduces the experimental results of chapters II and III, namely the modulation of feedback responses to perturbation according to the change in target structure. The second algorithm aims at navigating non-convex composition cost functions that can for instance be used when several targets are available. This algorithm integrated both low-level sensorimotor and high-level cognitive factors to predict the outcome of online decision processes. By applying this methodology to the tasks investigated in IV, we could reproduce the observed behavior and we could even generalize to other similar work. With this modeling work, we extend the current control models underlying human reaching movements and open new perspectives for future work.

2 Future work

2.1 Movement duration

The results we presented in this thesis as well as those reported in previous studies revealed that the duration of reaching movements is a highly variable kinematics parameter. Fitt's demonstrated for instance that movement duration varies for different accuracy demands [Fitts, 1954]. More recently, this parameter was shown to be dependent on factors such as the structure of the goal target [Knill et al., 2011, Nashed et al., 2012, Lowrey et al., 2017, De Comite et al., 2021] or the reward associated with the task [Shadmehr et al., 2010a, Haith et al., 2012]. However, whether this parameter is known by the agent at the beginning or even during movement is still unknown and will be an extremely interesting question to address in order to have further insight into the reaching control policy. Theoretically, we could identify the theoretical frameworks that consider that movement duration is known or unknown at the beginning of movement.

In the optimal feedback control framework proposed by Todorov [Todorov and Jordan, 2002], movement duration is assumed to be known during movement planning and will be one of the parameters that characterizes the control policy. Even if the feedback gains are not explicitly dependent on this parameter, they are implicitly dependent on movement duration as they are computed recursively backwards from the last time step[1]. This model has the disadvantage that it does not explain the increase in movement duration observed when mechanical [Nashed et al., 2012, De Comite et al., 2021] or visual [Dimitriou et al., 2013, Franklin et al., 2016] perturbation occur during movement.

Other theoretical frameworks assume that movement duration can be obtained from control policy itself. For instance Guigon and colleagues [Guigon et al., 2007] demonstrated that movement duration could be obtained from a desired level of effort. This idea was further developed by Rigoux and colleagues [Rigoux and Guigon, 2012] that integrated the idea in a framework that could model optimal decision making and motor control. An alternative to the LQG controller we presented before could also model movement duration as an consequence of the control policy. This can be done by using an infinite horizon optimal feedback controller [Phillis, 1985, Qian et al., 2013, Li et al., 2018] which computes the stabilizing solution of the Optimal Control problem.

An experimental study supporting any of the aforementioned hypotheses will be very interesting as it will allow to select more appropriately the theoretical framework used to model these reaching movements.

2.2 Online adjustments in control policy elicited by reward distribution

With the experiments of Chapter II and III we demonstrated that there exists a mechanism able to dynamically update the control policy used by humans to perform movement. More specifically, we modified the target structure, either suddenly or dynamically to investigate whether online changes in task demands could elicit online change in control policy. The task parameters we changed in these experiments were the target structure or the presence of obstacles but we do not know what

[1] At least in the finite horizon LQG formulation that we developed in this thesis

would happen if the factors that changed during movement were related to the reward. Indeed, our results and others [Manohar et al., 2019, Esteves et al., 2016, Summerside et al., 2018, Codol et al., 2020] demonstrated that target reward has an impact on the selection of the control policy. Moreover, the second experiment we performed in Chapter 4 and a recent study by Marti-Marca [Martí-Marca et al., 2020] showed that the reward distribution across the redundant dimension of the target or across the different targets had an impact on the online motor decisions.

An experiment probing the effect of online modification of reward distribution across target or across the redundant dimension of the target would be interesting because it will provide insight into the impact of reward on the feedback mechanism conveying task-related information and might eventually shed the light on the way reward is integrated into the selection of the control policy.

2.3 Mechanism underlying online change in control

An outstanding question is to understand what is the mechanism underlying these online change in control policy that we reported in Chapters 2 and 3. This mechanism is probably recruiting other circuits than the visuomotor reflex as supported by the difference in the response delays (150ms for the online adjustments in control policy [De Comite et al., 2021] versus 100ms for the visuomotor reflex [Georgopoulos et al., 1981]).

Since the online change in control policy we reported consists in planning and executing a new control policy, it probably recruits, amongst others, the neural circuits associated with the evaluation of cost and effort: the basal ganglia. To confirm this hypotheses, further study could for instance compare the ability of patients suffering from Parkison's disease to perform this task as it has been reported that these patients are more sensitive to motor cost [Mazzoni et al., 2007].

Another interesting question regarding the mechanism underlying these online adjustments in control policy would be whether it could somehow be related to the online learning mechanisms. Indeed, a recent study [Crevecoeur et al., 2020b] reported that human participants were able to adapt to new dynamics within the course of a trial in 250ms [Crevecoeur et al., 2020a]. Both mechanism requires more than reflex

responses as one involve a re-evaluation of the internal dynamic inherent to motor learning [Shadmehr and Mussa-Ivaldi, 1994] while the other one involves a re-evaluation of the control policy through motor planning [Wong et al., 2015]. However, it is still unknown if these two mechanisms could be carried out by the same neural pathways or whether they are two separates mechanisms.

2.4 Beyond Optimal Feedback Control

Despite being an extremely powerful model to simulate the control of reaching movements, Optimal Feedback Control (OFC) alone is not enough to explain the results reported in this thesis as well as some others presented in past studies. Indeed, OFC theory struggles to model reaching to redundant targets (that can be modeled for very specific redundancy such as rectangular targets [Nashed et al., 2012]), multiple targets (that requires an additional decision making component [Nashed et al., 2014], see the implementation in chapter V), online adjustments in control policy (necessitates an online re-computation of the optimal control policy, see [Lee, 2011] for a review on model predictive control and chapter V for an implementation), reward (that can difficulty modeled differently that target of different size) or even motor learning (that needs an additional learning process, see [Smith et al., 2006] for instance). The modeling of reaching movements is a primordial tool to understand their properties and our results demonstrate that the OFC framework might not be enough to grasp the complexity of the observed behaviors. A theoretical framework that could explain all these behaviors in addition to the features already explained by OFC would be an extremely interesting model to develop.

We identified deep reinforcement learning based control as a potential candidate to model reaching movements while tackling the problems mentioned above. The basic premise of the reinforcement learning framework is that an agent learns the ideal behavior in a given environment from experience collected while interacting with that environment [Sutton and Barto, 2018]. More precisely, the agent learns to associate an action to each state, which corresponds to the definition of control policy developed earlier. Until recently, reinforcement learning problems were confined to very specific tasks (grid worlds) because of algorithmic limitations but since the development of new algorithms leveraging the computational power of deep neural networks, the use

of reinforcement learning agents in the control field has increased. Indeed, the development of the DDPG algorithm by Lillicrap and colleagues [Lillicrap et al., 2019] allowed the modeling and the control of complex systems such as humanoids or robot arms [Levine et al., 2018, Levine et al., 2016, Schulman et al., 2016, Mnih et al., 2015].

Many reasons support this candidacy of deep reinforcement learning as a framework to explore for the modeling of reaching movements. First, on an neurophysiological point of view, reinforcement learning processes are used for a few decades in the neuroscience field as it was demonstrated that the brain had neural substrate supporting prediction and reward [Schultz et al., 1993, Schultz et al., 1997]. More closely related to the content of this thesis, the field of motor learning has shared ideas and terminology with reinforcement learning as demonstrated by the recent motor learning paper from Hadjiosif [Hadjiosif et al., 2021] which is reminiscent of the paper of Sutton on Policy gradient methods [Sutton, 2000]. Besides these experimental motivations, there is also a certain mathematical logic to expand Optimal Control with reinforcement learning for the specific model of upper limb reaching movement. Indeed, the theoretical framework underlying reinforcement learning is very similar to that of optimal control because, in both frameworks, the agent's actions are determined by the optimization of a performance criterion based on the interaction of the agent with an environment [Sutton and Barto, 2018]. The big differences between these frameworks resides in the way these optimizations are performed and the hypotheses that are assumed a fortiori.

Deep reinforcement learning algorithms based on policy optimization seem to be appropriate candidates to model these reaching movements and would require further investigations. The arguments developed in the previous paragraphs demonstrate that we are not trying to surf of the hype and that it makes sense to give it a try.

3 Concluding words

In this thesis, we investigate the imbrication of movement planning and execution. We first demonstrate that the control policies used during reaching movement are not frozen after movement planning and can be adjusted online if the task demands unexpectedly change. Elaborating on that, we then report that these online adjustments in control policy

result in fact from a continuous and continuous mechanism that conveys information about the task demands through a feedback loop. This additional feedback loop is sensitive to the rate of change of the task demands making it an exquisitely complex mechanism. Finally, we demonstrate that the robustness of the selected control policy influences its flexibility: more robust policy will be less flexible.

Together, these findings demonstrate the terrific imbrication of the planning and execution facets of sensorimotor control. We indeed demonstrated, through different experimental paradigms that planning can happen during execution which tears apart the strict dichotomy that could have existed before.

Bibliography

[Abe et al., 2011] Abe, M., Schambra, H., Wasserman, E. M., Luckenbaugh, D., Schweighofer, N., and Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology*, 21(7):557–562.

[Alexander, 1996] Alexander, R. M. (1996). *Optima for animals*. Princeton University Press.

[Alexander, 1997] Alexander, R. M. (1997). A minimum energy cost hypothesis for human arm trajectories. *Biological Cybernetics*, 76(2):97–105.

[Angelaki et al., 2009] Angelaki, D. E., Gu, Y., and DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current opinion in neurobiology*, 19(4):452–458.

[Astrom and Hagglund, 1984] Astrom, K. and Hagglund, T. (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, 20(5):645–651.

[Astrom and Hagglund, 1995] Astrom, K. J. and Hagglund, T. (1995). *PID controllers: theory, design, and tuning*.

[Atkeson and Hollerbach, 1985] Atkeson, C. G. and Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience*, 5(9):2318–2330.

[Avraham and Nisky, 2020] Avraham, C. and Nisky, I. (2020). The effect of tactile augmentation on manipulation and grip force control during force-field adaptation. *Journal of Neuroengineering and Rehabilitation*, 17(1).

[Aziz-Zadeh et al., 2004] Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., and Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *European journal of Neuroscience*, 19:2609–2612.

[Bahill and McDonald, 1983] Bahill, A. T. and McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research*, 23(12):1573–1583.

[Baraduc et al., 2013] Baraduc, P., Thobois, S., Gan, S., Broussolle, E., and Desmurget, M. (2013). A common optimization principle for motor execution in healthy subjects and parkinsonian patients. *Journal of Neuroscience*, 33(2):665–677.

[Barnes and Asselman, 1991] Barnes, G. R. and Asselman, P. T. (1991). The mechanism of prediction in human smooth pursuit eye movements. *Journal of Physiology*, 439:439–461.

[Barret and Glencross, 1989] Barret, N. and Glencross, D. J. (1989). Response amendments during manual aiming movements to double-steps targets. *Acta Psychologica*, 70(3):205–217.

[Basar and P, 1991] Basar, T. and P, B. (1991). *H infinity - Optimal control and related minimax design problems*. Birkhaüser, Boston, Massachussets.

[Bastian et al., 1996] Bastian, A. J., Martin, T. A., Keating, J. G., and Thach, W. T. (1996). Cerebellar ataxia: abnormal control of interaction torques across multiple joints. *Journal of Neurophysiology*, 76(1):492–509.

[Bedingham and Tatton, 1984] Bedingham, W. and Tatton, W. G. (1984). Dependence of EMG Responses Evoked by Imposed Wrist Displacements on Pre-existing Activity in the Stretched Muscles. (May):272–280.

[Benjamin and Berger, 2018] Benjamin, D. J. and Berger, J. O. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1):6–10. Publisher: Springer US.

[Bernstein, 1967] Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Oxford, New York, Pergamon Press.

[Berret et al., 2011] Berret, B., Chiovetto, E., Nori, F., and Pozzo, T. (2011). Manifold reaching paradigm: How do we handle target redundancy? *Journal of Neurophysiology*, 106(4):2086–2102.

[Berret and Jean, 2016] Berret, B. and Jean, F. (2016). Why Don ' t We Move Slower ? The Value of Time in the Neural Control of Action. 36(4):1056–1070.

[Bian et al., 2020] Bian, T., Wolpert, D. M., and Jiang, Z. P. (2020). Model-free robust optimal feedback mechanisms of biological motor control. *Neural Computation*, 32(3):562–595.

[Biess et al., 2007] Biess, A., Liebermann, D. G., and Flash, T. (2007). A computational model for redundant human three-dimensional pointing movements: Integration of independent spatial and temporal motor plans simplifies movement dynamics. *Journal of Neuroscience*, 27(48):13045–13064.

[Biguer et al., 1984] Biguer, B., Prablanc, C., and Jeannerod, M. (1984). The contribution of coordinated eye and head movements in hand pointing accuracy. *Experimental Brain Research*, 55:462–469.

[Blohm et al., 2020] Blohm, G., Kording, K. P., and Schrater, P. R. (2020). A how-to-model guide for neuroscience. *eNeuro*, 7(1).

[Blouin et al., 1995] Blouin, J., Gauthier, G. M., and Vercher, J. L. (1995). Internal representation of gaze direction with and without retinal inputs in man. *Neuroscience Letters*, 183(3):187–189.

[Bock, 1986] Bock, O. (1986). Contribution of retinal versus extraretinal signals towards visual localization in goal-directed movements. *Experimental Brain Research*, 64(3):476–482.

[Bock, 1993] Bock, O. (1993). Localization of objects in the peripheral visual field. *Behavioral Brain Research*, 56(1):77–84.

[Bogacz et al., 2010] Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., and Niewenhuis, S. (2010). The neural basis of speed-accuracy tradeoff. *Trends in Neuroscience*, 33(1):10–16.

[Boulinguez and Nougier, 1999] Boulinguez, P. and Nougier, V. (1999). Control of goal-directed movements: the contribution of orienting of visual attention and motor preparation. *Acta Psychologica*, 103(1):21–45.

[Brenner and Smeets, 1997] Brenner, E. and Smeets, J. B. (1997). Fast responses of the human hand to changes in target position. *Journal of Motor Behavior*, 29(4):297–310.

[Brooks et al., 2015] Brooks, J. X., Carriot, J., and Cullen, K. E. (2015). Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nature Neuroscience*, 18:1310–1317.

[Brotchie et al., 1995] Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375(6528):232–235.

[Burdet et al., 2001] Burdet, E., Osu, R., Franklin, D. W., Milner, T. E., and Kawato, M. (2001). The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, 414:446–449.

[Burk et al., 2014] Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., and Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PLOS ONE*, 9(3).

[Burke et al., 1976] Burke, D., Hagbarth, K. E., Löfstedt, L., and Wallin, B. G. (1976). The responses of human muscle spindle endings to vibration of non-contracting muscles. *The journal of Physiology*, 203(3):673–693.

[Calencie and Bawa, 1985] Calencie, B. and Bawa, P. (1985). Firing pattern of human flexor carpi radialis motor units during the stretch reflex. *Journal of Neurophysiology*, 53(5):1179–1193.

[Cameron et al., 2009] Cameron, B. D., Cressman, E. K., Franks, I. M., and R, C. (2009). Cognitive constraint on the "automatic pilot" for the hand: Movement intention influences the hand's susceptibility to involuntary online corrections. *Consciousness and Cognition*, 18(3):646–652.

[Cameron et al., 2014] Cameron, B. D., De la Malla, C., and Lopez-Martinez, J. (2014). The role of differential delays in integrating transient visual and proprioceptive information. *Frontiers in Psychology*, 5(50).

[Camponogara and Volcic, 2019] Camponogara, I. and Volcic, R. (2019). Grasping adjustments to haptic, visual, and visuo-haptic object perturbations are contingent on the sensory modality. *Journal of Neurophysiology*, 122(6):2614–2620.

[Capaday et al., 1994] Capaday, C., Forget, R., and Milner, T. (1994). A reexamination of the effect of instruction on the long latency stretch reflex response of the flexor pollicis longus muscle. *Experimental Brain Research*, 100(3):515–521.

[Capaday and Stein, 1986] Capaday, C. and Stein, R. B. (1986). Amplitude modulation of the soleus h-reflex in the human during walking and standing. *Journal of Neuroscience*, 6(5):1308–1313.

[Carlton and Carlton, 1987] Carlton, L. G. and Carlton, M. J. (1987). Response amendment latencies during discrete arm movements. *Journal of Motor Behavior*, 19(2):227–239.

[Carroll et al., 2019] Carroll, T. J., Mcnamee, D., Ingram, J. N., and Wolpert, D. M. (2019). Rapid visuomotor responses reflect value-based decisions. *Journal of Neuroscience*, 39(20):3906–3920.

[Castiello et al., 2006] Castiello, U., Zucco, G. M., Parma, V., Ansuini, C., and Tirindelli, R. (2006). Cross-modal interactions between olfaction and vision when grasping. *Chemical Senses*, 31(7):665–671.

[Cesonis and Franklin, 2021] Cesonis, J. and Franklin, D. (2021). Mixed-horizon optimal feedback control as a model of human movement. *arXiv*.

[Cesonis and Franklin, 2020] Cesonis, J. and Franklin, D. W. (2020). Time-to-target simplifies optimal control of visuomotor feedback responses. *eNeuro*, 7(2):514–519.

[Chapman et al., 2010] Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., and Goodale, M. A. (2010). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition*, 116(2):168–176.

[Chapman and Goodale, 2008] Chapman, C. S. and Goodale, M. A. (2008). Missing in action : the effect of obstacle position and size on avoidance while reaching. *Experimental Brain Research*, 191:83–97.

[Chapman et al., 2002] Chapman, H., Gavrilescu, M., Wang, H., Egan, G., and Castiello, U. (2002). Posterior parietal cortex control of reach-to-grasp movements in humans. *European Journal of Neuroscience*, 15(12).

[Cheney and Preston, 1976] Cheney, P. D. and Preston, J. B. (1976). Classification and response characteristics of muscle spindle afferents in the primate. *Journal of Neurophysiology*, 39(1):1–8.

[Churchland and Shenoy, 2007] Churchland, M. and Shenoy, K. V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *Journal of Neurophysiology*, 97(6):4235–4257.

[Churchland et al., 2012] Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487:51–56.

[Cisek, 2012] Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinion in Neurobiology*, 22(6):927–936.

[Cisek et al., 2009] Cisek, P., Puskas, G. A., and El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29(37):11560–11571.

[Cluff et al., 2015] Cluff, T., Crevecoeur, F., and Scott, S. H. (2015). A perspective on multisensory integration and rapid perturbation responses. *Vision Research*, 110:215–222.

[Cluff and Scott, 2015] Cluff, T. and Scott, S. H. (2015). Apparent and actual trajectory control depend on the behavioral context in upper limb motor tasks. *Journal of Neuroscience*, 35(36):12465–12476.

[Codol et al., 2020] Codol, O., Holland, P. J., Manohar, S. G., and Galea, J. M. (2020). Reward-based improvements in motor control are driven by multiple error-reducing mechanisms. *Journal of Neuroscience*, 40(18):3604–3620.

[Colebatch et al., 1979] Colebatch, J. G., Gandevia, S. C., McCloskey, D. I., and Potter, E. K. (1979). Subject instruction and long latency responses to muscle stretch. *Journal of Physiology*, 292:527–534.

[Connolly et al., 2009] Connolly, P. M., Bennur, S., and Gold, J. I. (2009). Correlates of perceptual learning in an oculomotor decision variable. *Journal of Neuroscience*, 29(7):2136–2150.

[Cook et al., 2019] Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y., Bloniarz, A. E., Yakolev, M. A., Nguyen, K. C. Q., Tang, L. T.-H., Bayer, E. A., Duerr, J. S., Bülow, H. E., Hobert, O., Hall, D. H., and Emmons, S. W. (2019). Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571:63–71.

[Cos et al., 2011] Cos, I., Bélanger, N., and Cisek, P. (2011). The influence of predicted arm biomechanics on decision making. *Journal of Neurophysiology*, 105(6):3022–3033.

[Cos et al., 2014] Cos, I., Duque, J., and Cisek, P. (2014). Rapid prediction of biomechanical costs during action decisions. *Journal of neurophysiology*, 112(6):1256–1266.

[Cos et al., 2021] Cos, I., Pezzulo, G., and Cisek, P. (2021). Changes of mind after movement onset: a motor-state dependent decision-making process. *bioRxiv*, page 6.

[Crago et al., 1976] Crago, P. E., Houk, J. C., and Hasan, Z. (1976). Regulatory actions of human stretch reflex. *Journal of Neurophysiology*, 39(5):925–935.

[Crevecoeur et al., 2013] Crevecoeur, F., Kurtzer, I., Bourke, T., and Scott, S. H. (2013). Feedback responses rapidly scale with the urgency to correct for external perturbations. *Journal of Neurophysiology*, 110(6):1323–1332.

[Crevecoeur et al., 2014] Crevecoeur, F., Kurtzer, I., and Scott, S. H. (2014). Fast corrective responses are evoked by perturbations approaching the natural variability of posture and movement tasks. *Journal of Neurophysiology*, 107(10):2821–2832.

[Crevecoeur et al., 2020a] Crevecoeur, F., Mathew, J., Bastin, M., and Lefèvre, P. (2020a). Feedback adaptation to unpredictable force fields in 250ms. *eNeuro*, 7(2).

[Crevecoeur et al., 2016] Crevecoeur, F., Munoz, D. P., and Scott, S. H. (2016). Dynamic Multisensory Integration: Somatosensory Speed Trumps Visual Accuracy during Feedback Control. *The Journal of Neuroscience*, 36(33):8598–8611.

[Crevecoeur and Scott, 2014] Crevecoeur, F. and Scott, S. H. (2014). Beyond Muscles Stiffness: Importance of State-Estimation to Account for Very Fast Motor Corrections. *PLoS Computational Biology*, 10(10).

[Crevecoeur et al., 2019] Crevecoeur, F., Scott, S. H., and Cluff, T. (2019). Robust control in human reaching movements: A model-free strategy to compensate for unpredictable disturbances. *Journal of Neuroscience*, 39(41):8135–8148.

[Crevecoeur et al., 2011] Crevecoeur, F., Sepulchre, R. J., Thonnard, J. L., and Lefèvre, P. (2011). Improving the state estimation for optimal control of stochastic processes subject to multiplicative noise. *Automatica*, 47(3):591–596.

[Crevecoeur et al., 2020b] Crevecoeur, F., Thonnard, J.-L., and Lefèvre, P. (2020b). A very fast time scale of human motor adaptation: within movement adjustments of internal representations during reaching. *eNeuro*, 7(1).

[Cross et al., 2019] Cross, K. P., Cluff, T., Takei, T., and Scott, S. H. (2019). Visual feedback processing of the limb involves two distinct phases. *Journal of Neuroscience*, 39(34):6751–6765.

[Datta et al., 2019] Datta, S. R., Anderson, D. J., Branson, K., Perona, P., and Leifer, A. (2019). Computational neuroethology : A call to action. *Neuron*, 104(1):11–24.

[Day and Lyon, 2000] Day, B. L. and Lyon, I. N. (2000). Voluntary modification of automatic arm movements evoked by motion of a visual target. *Experimental Brain Research*, 130(2):159–168.

[De Comite et al., 2021] De Comite, A., Crevecoeur, F., and Lefèvre, P. (2021). Online modification of goal-directed control in human reaching movements. *Journal of Neurophysiology*, 125(5):1883–1898.

[Delhaye et al., 2018] Delhaye, B. P., Long, K. H., and Bensmaia, S. J. (2018). Neural basis of touch and proprioception in primate cortex. *Comprehensive Physiology*, 8(4):1575–1602.

[Delp et al., 2007] Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., and Thelen, D. G. (2007). Opensim: Open-source software to create an analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11):1940–1950.

[Desmurget et al., 1999] Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., and Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nature Neuroscience*, 2(6):563–567.

[Desmurget and Grafton, 2000] Desmurget, M. and Grafton, S. T. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11):423–431.

[Desmurget et al., 1998] Desmurget, M., Pélisson, D., Rossetti, Y., and Prablanc, C. (1998). From eye to hand: Planning goal-directed movements. *Neuroscience & Biobehavioral Reviews*, 22(6):761–788.

[Desmurget et al., 1997] Desmurget, M., Rossetti, Y., Jordan, M., Meckler, C., and Prablanc, C. (1997). Viewing the hand prior to movement improves accuracy of pointing performed toward the unseen contralateral hand. *Experimental Brain Research*, 115:180–186.

[Diedrichsen, 2007] Diedrichsen, J. (2007). Optimal task-dependent changes of bimanual feedback control and adaptation. *Current Biology*, 17(19):1675–1679.

[Diedrichsen et al., 2010] Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2010). The coordination of movement: optimal feedback control and beyond. *Trends in Cognitive Sciences*, 14(1):31–39.

[Dimitriou and Edin, 2008a] Dimitriou, M. and Edin, B. B. (2008a). Discharges in human muscle receptor afferents during block grasping. *Journal of Neuroscience*, 48(48):12632–12642.

[Dimitriou and Edin, 2008b] Dimitriou, M. and Edin, B. B. (2008b). Discharges in human muscle spindle afferents during a key-pressing task. *Journal of Physiology*, 586(22):5455–5470.

[Dimitriou et al., 2013] Dimitriou, M., Wolpert, D. M., and Franklin, D. W. (2013). The Temporal Evolution of Feedback Gains Rapidly Update to Task Demands. *Journal of Neuroscience*, 33(26):10898–10909.

[Ding and Gold, 2012] Ding, L. and Gold, J. I. (2012). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, 22:1052–1067.

[Doyle and Walker, 2002] Doyle, M. C. and Walker, R. (2002). Multisensory interactions in saccade target selection: Curved saccade trajectories. *Experimental Brain Research*, 142:116–130.

[Driver and Spence, 1998] Driver, J. and Spence, C. (1998). Crossmodal attention. *Current opinion in Neurobiology*, 8(2):245–253.

[Dudman and Krakauer, 2016] Dudman, J. T. and Krakauer, J. W. (2016). The basal ganglia: from motor commands to the control of vigor. *Current Opinion in Neurobiology*, 37:158–166.

[Dufresne et al., 1980] Dufresne, J. R., Soechting, J. F., and Terzuolo, C. A. (1980). Modulation of the myotatic reflex gain in man during intentional movements. *Brain Research*, 193(1):67–84.

[Duysens et al., 1993] Duysens, J., Tax, A. A., Trippel, M., and Dietz, V. (1993). Increased amplitude of cutaneous reflexes during human running as compared to standing. *Brain Research*, 613(2):230–238.

[Ebner and Pasalar, 2008] Ebner, T. J. and Pasalar, S. (2008). Cerebellum predicts the future motor state. *Cerebellum*, 7(4):583–588.

[Edin and Vallbo, 1990a] Edin, B. and Vallbo, A. (1990a). Dynamic response of human muscle spindle afferents to stretch. *Journal of Neurophysiology*, 63(6):1297–1306.

[Edin and Vallbo, 1990b] Edin, B. B. and Vallbo, A. B. (1990b). Muscle afferents responses to isometric contractions and relaxations in humans. *Journal of Neurophysiology*, 63(6):1307–131.

[Efron, 1979] Efron, B. (1979). Bootstrap methods: Another look at the jacknife. *The Annals of Statistics*, 7(1):1–26.

[Enachescu et al., 2021] Enachescu, V., Schrater, P., Schaal, S., and Christopoulos, V. (2021). Action planning and control under uncertainty emerge through a desirability-driven competition between parallel encoding motor plans. *PLoS Computational Biology*, 17(10):e1009429.

[Ernst and Banks, 2002] Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433.

[Esteves et al., 2016] Esteves, P. O., Oliveira, L. A., Nogueira-Campos, A. A., Saunier, G., Pozzo, T., Oliveira, J. M., Rodrigues, E. C., Volchan, E., and Vargas, C. D. (2016). Motor planning of goal-directed action is tuned by the emotional valence of the stimulus: A kinematic study. *Scientific Reports*, 6:1–7.

[Fitts, 1954] Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381391.

[Flash, 1987] Flash, T. (1987). The control of hand equilibrium trajectories in multi-joint arm movements. *Biological Cybernetics*, 57:257–274.

[Flash and Hogan, 1985] Flash, T. and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7):1688–1703.

[Forgaard et al., 2021] Forgaard, C., Reschechtko, S., Gribble, P. L., and Pruszynski, J. A. (2021). Skin and muscle receptors shape coordinated fast feedback responses in the upper limb. *Current Opinion in Physiology*, 20:198–205.

[Franklin et al., 2016] Franklin, D. W., Reichenbach, A., Franklin, S., and Diedrichsen, J. (2016). Temporal Evolution of Spatial Computations for Visuomotor Control. *Journal of Neuroscience*, 36(8):2329–2341.

[Franklin and Wolpert, 2008] Franklin, D. W. and Wolpert, D. M. (2008). Specificity of Reflex Adaptation for Task-Relevant Variability. *Journal of Neuroscience*, 28(52):14165–14175. ISBN: 0270-6474.

[Galea et al., 2015] Galea, J. M., Mallia, E., Rothwell, J., and Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature Neuroscience*, 18:597–602.

[Gallivan et al., 2016a] Gallivan, J. P., Bowman, N. A. R., Chapman, C. S., Wolpert, D. M., and Flanagan, J. R. (2016a). The sequential encoding of competing action goals involves dynamic restructuring of motor plans in working memory. *Journal of Neurophysiology*, 115(6):3113–3122.

[Gallivan et al., 2016b] Gallivan, J. P., Logan, L., Wolpert, D. M., and Flanagan, J. R. (2016b). Parallel specification of competing sensorimotor control policies for alternative action options. *Nature Neuroscience*, 19(2):320–326.

[Gallivan et al., 2017] Gallivan, J. P., Stewart, B. M., Baugh, L. A., Wolpert, D. M., and Flanagan, J. R. (2017). Rapid automatic motor encoding of competing reach options. *Cell Report*, 18(7):1619–1626.

[Gazzola et al., 2006] Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, 16:1824–1829.

[Georgopoulos et al., 1981] Georgopoulos, A. P., Kalaska, J. F., and Massey, J. T. (1981). Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty, and change in target location. *Journal of Neurophysiology*, 46(4):725–743.

[Glazebrook, 2018] Glazebrook, C. M. (2018). Catching the Integration Train : A Look Into the Next 10 Years of Motor-Control and Motor-Learning Research. *Kinesiology Review*, 7(2):130–141.

[Gold and Shadlen, 2000] Gold, J. I. and Shadlen, M. N. (2000). Representation of a perceptual decision in developping oculomotor commands. *Nature*, 404:390–394.

[Gold and Shadlen, 2003] Gold, J. I. and Shadlen, M. N. (2003). The influence of behavioral context on the representation of a perceptual decision in developping oculomotor commands. *Journal of Neuroscience*, 23(2):632–651.

[Gold and Shadlen, 2007] Gold, J. I. and Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, 30:535–574.

[Gomi and Kawato, 1996] Gomi, H. and Kawato, M. (1996). Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science*, 272(5258):117–120.

[Gomi and Kawato, 1997] Gomi, H. and Kawato, M. (1997). Human arm stiffness and equilibrium-point trajectory during multi-joint movement. *Biological Cybernetics*, 76:163–171.

[Goodale et al., 1986] Goodale, M. A., Pelisson, D., and 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.

[Goodwin et al., 1972] Goodwin, G. M., McCloskey, D. I., and Matthews, P. B. (1972). The contribution of muscle afferents to kinaesthesia shown by vibration induced illusion of movement and by the effects of paralysing joint afferents. *Brain*, 95:705–748.

[Gordon and Ghez, 1987] Gordon, J. and Ghez, C. (1987). Trajectory control in targeted force impulses. iii. compensatory adjustments for initial errors. *Experimental Brain Research*, 67(2):253–269.

[Gordon et al., 1994] Gordon, J., Ghilardi, M., Cooper, S. E., and Ghez, C. (1994). Accuracy of planar reaching movements. *Experimental Brain Research*, 99:112–130.

[Gréa et al., 2002] Gréa, H., Pisella, L., Rosetti, Y., Desmurget, M., Tilikete, C., Grafton, S., Prablanc, C., and Vighetto, A. (2002). A lesion of the posterior parietal cortex disrupts on-line adjustments during aiming movements. *Neuropsychologica*, 40(13):2471–2480.

[Graziano and Gross, 1993] Graziano, M. S. A. and Gross, C. G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive field. *Experimental Brain Research*, 97:96–109.

[Gritsenko and Kalaska, 2010] Gritsenko, V. and Kalaska, J. F. (2010). Rapid online correction is selectively suppressed during movement with a visuomotor transformation. *Journal of Neurophysiology*, 104(6):3084–3104.

[Gritsenko et al., 2009] Gritsenko, V., Yakovenko, S., and Kalaska, J. F. (2009). Integration of predictive feedforward and sensory feedback signals for online control of visually guided movement. *Journal of Neurophysiology*, 102(2):914–930.

[Guigon et al., 2007] Guigon, E., Baraduc, P., and Desmurget, M. (2007). Computational motor control: Redundancy and invariance. *Journal of Neurophysiology*, 97(1):331–347.

[Hadjiosif et al., 2021] Hadjiosif, A. M., Krakauer, J. W., and Haith, A. M. (2021). Did we get sensorimotor adaptation wrong? implicit adaptation as direct policy updating rather than forward-model-based learning. *Journal of Neuroscience*, 41(12):2747–2761.

[Haggard et al., 1995] Haggard, P., Hutchinson, K., and Stein, J. (1995). Patterns of coordinated multi-joint movement. *Experimental Brain Research*, 107(2):254–266.

[Haith et al., 2016] Haith, A. M., Pakpoor, J., and Krakauer, J. W. (2016). Independence of movement preparation and movement initiation. *Journal of Neuroscience*, 36(10):3007–3015.

[Haith et al., 2012] Haith, A. M., Reppert, T. R., and Shadmehr, R. (2012). Evidence for hyperbolic temporal discounting of reward in control of movements. *Journal of Neuroscience*, 32(34):11727–11736.

[Hammond, 1956] Hammond, P. (1956). The influence of prior instruction to the subject on an apparently involuntary neuro-muscular response. *Journal of Physiology*, 132(1):17–8P.

[Hanes and Schall, 1996] Hanes, D. P. and Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274(5286):427–430.

[Hanks et al., 2014] Hanks, T., Kiani, R., and Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area lip. *eLife*, 3:e02260.

[Harris and Wolpert, 1998] Harris, C. M. and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–784.

[Heitz, 2014] Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8:150.

[Hernandez-Castillo et al., 2020] Hernandez-Castillo, C. R., Maeda, R. S., Pruszynski, J. A., and Diedrichsen, J. (2020). Sensory information from a slipping object elicits a rapid and automatic shoulder response. *Journal of Neurophysiology*, 123(3):1103–1112.

[Herter et al., 2009] Herter, T. M., Korbel, T., and Scott, S. H. (2009). Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. *Journal of Neurophysiology*, 101(1):150–163.

[Herter et al., 2006] Herter, T. M., Kurtzer, I., Cabel, W., Haunts, K. A., and Scott, S. H. (2006). Characterization of torque-related activity in primary motor cortex during a multijoint postural task. *Journal of Neurophysiology*, 97(4):2887–2899.

[Hollerbach and Flash, 1982] Hollerbach, J. M. and Flash, T. (1982). Dynamic interactions between limb segments during planar arm movement. *Biological Cybernetics*, 44:67–77.

[Horwitz et al., 2004] Horwitz, G. D., Batista, A. P., and Newsome, W. T. (2004). Representation of an abstract perceptual decision in macaque superior colliculus. *Journal of Neurophysiology*, 91(5):2281–2296.

[Horwitz and Newsome, 1999] Horwitz, G. D. and Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, 284(5417):1158–1161.

[Howard and Tipper, 1997] Howard, L. A. and Tipper, S. P. (1997). Hand deviations away from visual cues : indirect evidence for inhibition. *Experimental Brain Research*, 113(1):144–152.

[Huang et al., 2012] Huang, H. J., Kram, R., and Ahmed, A. A. (2012). Reduction of metabolic cost during motor learning of arm reaching dynamics. *Journal of Neuroscience*, 32(6):2182–2190.

[Hulliger, 1984] Hulliger, M. (1984). The mammalian muscle spindle and its central control. *Reviews of Physiology, Biochemistry and Pharmacology*, 101.

[Ings and Chittka, 2008] Ings, T. C. and Chittka, L. (2008). Speed-accuracy tradeoffs and false alarms in bee responses to cryptic predators. *Current Biology*, 18:1520–1524.

[Izawa and Shadmehr, 2008] Izawa, J. and Shadmehr, R. (2008). On-line processing of uncertain information in visuomotor control. *Journal of Neuroscience*, 28(44):11360–11368.

[Jenmalm et al., 2000] Jenmalm, P., Dahlstedt, S., and Johansson, R. S. (2000). Visual and tactile information about object-curvature control fingertip forces and grasp kinematics in human dexterous manipulation. *Journal of Neurophysiology*, 84:2984–2997.

[Johansson and Westling, 1984] Johansson, R. S. and Westling, G. (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Experimental Brain Research*, 56:550–564.

[Johansson and Westling, 1988] Johansson, R. S. and Westling, G. (1988). Programmed and triggered actions to rapid load changes during precision grip. *Experimental Brain Research*, 71(1):72–86.

[Johnson, 2001] Johnson, K. O. (2001). The roles and functions of cutaneous mechanoreceptors. *Current Opinion in Neurobiology*, 11(4):455–461.

[Johnson and Hsiao, 1992] Johnson, K. O. and Hsiao, S. S. (1992). Neural mechanisms of tactual form and texture perception. *Annual Review Neuroscience*, 15:227–250.

[Johnson et al., 2000] Johnson, K. O., Yoshioka, T., and Vega-Bermudez, F. (2000). Tactile functions of mechanoreceptive afferents innervating the hand. *Journal of Clinical Neurophysiology*, 17(6):539–558.

[Kalaska, 2009] Kalaska, J. F. (2009). From intention to action: motor cortex and the control of reaching movements. *Progress in Motor Control*, 629:139–178.

[Kalaska et al., 1997] Kalaska, J. F., Scott, S. H., Cisek, P., and Sergio, L. E. (1997). Cortical control of reaching movements. *Current Opinion in Neurobiology*, 7(6):849–859.

[Kaminski and Gentile, 1989] Kaminski, T. R. and Gentile, A. M. (1989). A kinematic comparison of single and multijoint pointing movements. *Experimental Brain Research*, 78:547–556.

[Kasuga et al., 2021] Kasuga, S., Crevecoeur, F., Cross, K. P., Balalaie, P., and Scott, S. H. (2021). Integration of proprioceptive and visual feedback during online control of reaching. *Journal of Neurophysiology*, 0(0):null.

[Kaufman et al., 2014] Kaufman, M. T., Churchland, M. M., Ryu, S. I., and Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nature Neuroscience*, 17(3):440–448.

[Kawagoe et al., 1998] Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience*, 1:411–416.

[Kim and Shadlen, 1999] Kim, J. N. and Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, 2(2):176–185.

[Kimura et al., 2006] Kimura, T., Haggard, P., and Gomi, H. (2006). Transcranial magnetic stimulation over sensorimotor cortex disrupts anticipatory reflex gain modulation for skilled action. *Journal of Neuroscience*, 26(36):9272–9281.

[Kirk, 2004] Kirk, D. E. (2004). *Optimal control theory, an introduction*. Dover Publications, Inc.

[Klatzky and Lederman,] Klatzky, L. and Lederman, S. J. The intelligent hand. In Bower, G. H., editor, *Psychology of learning and motivation*, pages 121–151. San Diego : Academic Press.

[Klatzky et al., 1987] Klatzky, L. R., Lederman, S. J., and Reed, C. (1987). There's more to touch that meets the eye: the salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology*, 116:356–369.

[Klatzy et al., 2000] Klatzy, L. R., Pai, D. K., and Krotkov, E. P. (2000). Perception of materials from contact sounds. *Presence Teleoperators Virtual Environments*, 9:399–400.

[Knill et al., 2011] Knill, D. C., Bondada, A., and Chhabra, M. (2011). Flexible , Task-Dependent Use of Sensory Feedback to Control Hand Movements. *Journal of Neuroscience*, 31(4):1219–1237.

[Koivo and Tantuu, 1991] Koivo, H. N. and Tantuu, J. T. (1991). Tuning of pid controllers: survey of siso and mimo techniques. *Intelligent tuning and adaptive control*, pages 75–80.

[Kolossiatis et al., 2016] Kolossiatis, M., Charalambous, T., and Burdet, E. (2016). How variability and effort determine coordination at large forces. *Plos One*, 11(3):e0149512.

[Kording et al., 2007] Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PloS one*, 2(9):e943.

[Krakauer et al., 2017] Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3):480–490.

[Krakauer et al., 1999] Krakauer, J. W., Ghilardi, M.-F., and Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nature Neuroscience*, 2:1026–1031.

[Krakauer et al., 2000] Krakauer, J. W., Pine, Z. M., Ghilardin, M.-F., and Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, 20(23):8916–8924.

[Körding and Wolpert, 2004] Körding, K. P. and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247.

[Kurtzer et al., 2020] Kurtzer, I., Muraoka, T., Singh, T., Prasad, M., Chauhan, R., and Adhami, E. (2020). Reaching movements are automatically redirected to nearby options during target split. *Journal of Neurophysiology*, 124(4):10313–1028.

[Kurtzer et al., 2008] Kurtzer, I. L., Pruszynski, J. A., and Scott, S. H. (2008). Long-Latency Reflexes of the Human Arm Reflect an Internal Model of Limb Dynamics. *Current Biology*, 18(6):449–453.

[Kustov and Robinson, 1996] Kustov, A. A. and Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384:74–77.

[Kwon et al., 1982] Kwon, W. H., Bruckstein, A. M., and Kailath, T. (1982). Stabilizing state-feedback design via the moving horizon method. In *1982 21st IEEE Conference on Decision and Control*, pages 234–239.

[Lackner and Dizio, 1994] Lackner, J. R. and Dizio, P. (1994). Rapid adaptation to coriolis force perturbations of arm trajectory. *Journal of Neurophysiology*, 72(1):299–313.

[Lacquaniti and Soechting, 1984] Lacquaniti, F. and Soechting, J. F. (1984). Behavior of the stretch reflex in a multi-jointed limb. *Brain Research*, 311(1):161–166.

[Lacquaniti et al., 1986] Lacquaniti, F., Soechting, J.-F., and Terzuolo, S. A. (1986). Path constraints on point-to-point arm movements in three-dimensional space. *Neuroscience*, 17(2):313–324.

[Lacquaniti et al., 1983] Lacquaniti, F., Terzuolo, C., and Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54:115–130.

[Lakens, 2013] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(NOV):1–12.

[Laming, 1968] Laming, D. (1968). Information theory of choice-reaction times. *Behavioral Science*, 14(4):330–333.

[Lee, 2011] Lee, J. H. (2011). Model predictive control: Review of the three decades of development. *International Journal of Control, Automation and Systems*, 9(3):415–424.

[Levine et al., 2016] Levine, S., Finn, C., Darrell, T., and Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17:1–40.

[Levine et al., 2018] Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., and Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Journal of Robotics*, 37(4-5).

[Li and Todorov, 2004] Li, W. and Todorov, E. (2004). Iterative linear quadratic regulator design for nonlinear biological movement systems. In *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics*, pages 222–229.

[Li et al., 2018] Li, Z., Mazzoni, P., Song, S., and Qian, N. (2018). A single, continuously applied control policy for modeling reaching movements with and without perturbation. *Neural Computation*, 30(2):397–427.

[Lillicrap et al., 2019] Lillicrap, T. P., Hunt, J. J., Pritze, A., Heess, N., Erez, T., Y, T., Silver, D., and D, W. (2019). Continuous control with deep reinforcement learning. *arXiv*.

[Liu and Todorov, 2007] Liu, D. and Todorov, E. (2007). Evidence for the Flexible Sensorimotor Strategies Predicted by Optimal Feedback Control. *Journal of Neuroscience*, 27(35):9354–9368.

[Lowrey et al., 2017] Lowrey, C. R., Nashed, J. Y., and Scott, S. H. (2017). Rapid and flexible whole body postural responses are evoked from perturbations to the upper limb during goal-directed reaching. *Journal of Neurophysiology*, 117(3):1070–1083.

[Ma et al., 1994] Ma, B., Giat, Y., and Levine, W. S. (1994). The optimal control of a movement of the human upper extremity. *IFAC Proceedings*, 27:455–460.

[Ma-Wyatt and McKee, 2007] Ma-Wyatt, A. and McKee, S. P. (2007). Visual information throughout a reach determines endpoint precision. *Experimental Brain Research*, 179:55–64.

[Maeda et al., 2018] Maeda, R. S., Cluff, T., Gribble, P., and Pruszynski, J. A. (2018). Feedforward and feedback control share an internal model of the arm's dynamics. *Journal of Neuroscience*, 38(49):10505–10514.

[Maeda et al., 2017] Maeda, R. S., Cluff, T., Gribble, P. L., and Pruszynski, J. A. (2017). Compensating for intersegmental dynamics across the shoulder, elbow and wrist joints during feedforward and feedback control. *Journal of Neurophysiology*, 118(4):1984–1997.

[Maeda et al., 2020] Maeda, R. S., Kersten, R., and Pruszynski, J. A. (2020). Shared internal models for feedforward and feedback control of arm dynamics in non-human primates. *European Journal of Neuroscience*, 53(5):1605–1620.

[Manohar et al., 2015] Manohar, S. G., Chong, T. T., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., and Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, 25(13):1707–1716.

[Manohar et al., 2017] Manohar, S. G., Finzi, R. D., Drew, D., and Husain, M. (2017). Distinct Motivational Effects of Contingent and Noncontingent Rewards. *Psychological Science*, 28(7):1016–1026.

[Manohar et al., 2019] Manohar, S. G., Muhammed, K., Fallon, S. J., and Husain, M. (2019). Motivation dynamically increases noise resistane by internal feedback during movement. *Neuropsychologia*, 123(4):19–29.

[Marsden et al., 1976] Marsden, C. D., Merton, P. A., and Morton, H. B. (1976). Servo action in the human thumb. *Journal of Physiology*, 257(1):1–44.

[Martí-Marca et al., 2020] Martí-Marca, A., Deco, G., and Cos, I. (2020). Visual-reward driven changes of movement during action execution. *Scientific Reports*, 10(15527).

[Mathew et al., 2020] Mathew, J., Lefèvre, P., and Crevecoeur, F. (2020). Rapid changes in movement representations during human reaching could be preserved in memory for at least 850ms. *eNeuro*, 7(6).

[Matsumoto and Hikosaka, 2007] Matsumoto, M. and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, 447(7148):1111–1115.

[Matthews, 1933] Matthews, B. H. C. (1933). Nerve endings in mammalian muscle. *Journal of Physiology*, 78(1):1–53.

[Matthews, 1986] Matthews, P. B. (1986). Observations on the automatic compensation of reflex gain on varying the pre-existing level of motor discharge in man. *Journal of Physiology*, 374(May):73–90.

[Matthews, 1991] Matthews, P. B. (1991). The human stretch reflex and the motor cortex. *Trends in Neurosciences*, 14(3):87–91.

[Matthews, 1972] Matthews, P. B. C. (1972). *Mammalian muscle receptors and their central actions*. Edward Arnold.

[Mazzoni et al., 2007] Mazzoni, P., Hristova, A., and Krakauer, J. W. (2007). Why don't we move faster ? parkinson's disease, movement vigor and implicit motivation. *Journal of Neuroscience*, 27(27):7105–7116.

[Mazzoni and Krakaueur, 2006] Mazzoni, P. and Krakaueur, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *Journal of Neuroscience*, 26(14):3642–3645.

[McMillan, 1983] McMillan, G. (1983). *Tuning and control loop performance*. Momentum Press.

[McNamee and Wolpert, 2019] McNamee, D. and Wolpert, D. M. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:339–364.

[Megaw, 1974] Megaw, E. D. (1974). Possible modification to a rapid on-going programmed manual response. *Brain Research*, 71(2):425–441.

[Michalski et al., 2020] Michalski, J., Green, A. M., and Cisek, P. (2020). Reaching decisions during ongoing movements. *Journal of Neurophysiology*, 123(3):1090–1102.

[Mileusnic and Loeb, 2009] Mileusnic, M. P. and Loeb, G. E. (2009). Force estimation from ensembles of golgi tendon organs. *Journal of Neural Engineering*, 6.

[Mnih et al., 2015] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wiestra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529–533.

[Morasso, 1981] Morasso, P. (1981). Spatial control of arm movements. *Experimental Brain Research*, 42:223–227.

[Morel et al., 2017] Morel, P., Ulbrich, P., and Gail, A. (2017). What makes a reach movement effortful ? physical effort discounting supports common minimization principles in decision making and motor control. *PLOS Biology*, 15(6):1–23.

[Mortimer et al., 1981] Mortimer, J. A., Webster, D. D., and Dukich, T. A. (1981). Changes in short and long latency stretch responses during the transition from posture to movement. *Brain Research*, 229(2):337–351.

[Mosberger et al., 2016] Mosberger, A. C., de Clauser, L., Kasper, H., and Schwab, M. E. (2016). Motivational state, reward value, and pavlovian cues differentially affect skilled forelimb grasping in rats. *Learning & Memory*, 23:289–302.

[Mott and Sherrington, 1894] Mott, F. W. and Sherrington, C. S. (1894). Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. *Proceedings of the Royal Society of London*, 57:481–488.

[Muller and Dichgans, 1994] Muller, F. and Dichgans, J. (1994). Dyscoordination of pinch and lift forces during grasp in patients with cerebellar lesions. *Experimental Brain Research*, 101:485–492.

[Mutha et al., 2008] Mutha, P. K., Boulinguez, P., and Sainburg, R. L. (2008). Visual modulation of proprioceptive reflexes during movement. *Brain Research*, 1246:54–69.

[Nashed et al., 2012] Nashed, J. Y., Crevecoeur, F., and Scott, S. H. (2012). Influence of the behavioral goal and environmental obstacles on rapid feedback responses. *Journal of Neurophysiology*, 108(4):999–1009. ISBN: 1522-1598; 0022-3077.

[Nashed et al., 2014] Nashed, J. Y., Crevecoeur, F., and Scott, S. H. (2014). Rapid online selection between multiple motor plans. *Journal of Neuroscience*, 34(5):1769–1780.

[Nelson, 1983] Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46(2):135–147.

[Omrani et al., 2013] Omrani, M., Diedrichsen, J., and Scott, S. H. (2013). Rapid feedback corrections during a bimanual postural task. *Journal of Neurophysiology*, 109(1):147–161.

[Omrani et al., 2016] Omrani, M., Murnaghan, C. D., Pruszynski, J. A., and Scott, S. H. (2016). Distributed task-specific processing of somatosensory feedback for voluntary motor control. *eLife*, 5:1–17.

[Opris et al., 2011] Opris, I., Lebedev, M., and Nelson, R. J. (2011). Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity. *Frontiers in Neuroscience*, 5(61).

[Orban de Xivry et al., 2017] Orban de Xivry, J.-J., Legrain, V., and Lefèvre, P. (2017). Overlap of movement planning and movement execution reduces reaction time. *Journal of Neurophysiology*, 117(1):117–122.

[Orban de Xivry et al., 2008] Orban de Xivry, J. J., Missal, M., and Lefèvre, M. (2008). A dynamic representation of target motion drives predictive smooth pursuit during target blanking. *Journal of Vision*, 8(15):1–13.

[Paillard and Amblard, 1985] Paillard, J. and Amblard, B. (1985). Static versus kinetic visual cues for the processing of spatial relationships. In Ingle, D. J., Jeannerod, M., and Lee, D. N., editors, *Brain mechanisms and spatial vision*, pages 299–330. Springer.

[Palmer et al., 2005] Palmer, J., Huk, A. C., and Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5:376–404.

[Patchay et al., 2003] Patchay, S., Castiello, U., and Haggard, P. (2003). A cross-modal interference in grasping objects. *Psychonomic Bulletin and Review*, 10:924–931.

[Patchay et al., 2006] Patchay, S., Haggard, P., and Castiello, U. (2006). An object-centered reference frame for control of grasping: effects of grasping a distractor object on visuomotor control. *Experimental Brain Research*, 170:532–542.

[Peacock, 1983] Peacock, J. A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202:615–627.

[Phillis, 1985] Phillis, Y. A. (1985). Controller Design of Systems with Multiplicative Noise. (10):1017–1019.

[Pierrot-Deseillignly and Burke, 2005] Pierrot-Deseillignly, E. and Burke, D. (2005). *The circuitry of the human spinal cord: its role in motor control and movement disorders*. Cambridge University Press.

[Pélisson et al., 1986] Pélisson, D., Prablanc, C., Goodale, M. A., and Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb - II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Experimental Brain Research*, 62(2):303–311.

[Poscente et al., 2021] Poscente, S. V., Peters, R. M., Cashaback, J. G. A., and Cluff, T. (2021). Rapid feedback responses parallel the urgency of voluntary reaching movements. *Neuroscience*, 475:163–184.

[Prablanc et al., 1979a] Prablanc, C., Echallier, J. F., Komilis, E., and Jeannerod, M. (1979a). Optimal response of eye and hand motor systems in pointing at visual target - ii: Static and dynamic visual cues in the control of hand movement. *Biological Cybernetics*, 35:113–124.

[Prablanc et al., 1979b] Prablanc, C., Echallier, J. F., Komilis, E., and Jeannerod, M. (1979b). Optimal response of eye and hand motor systems in pointing to a visual target. *Biological cybernetics*, 35:113–124.

[Proske and Gandevia, 2009] Proske, U. and Gandevia, S. C. (2009). The kinaesthetic senses. *The journal of Physiology*, 587(17):4139–4146.

[Pruszynski et al., 2010] Pruszynski, J. A., King, G. L., Boisse, L., Scott, S. H., Flanagan, J. R., and Munoz, D. P. (2010). Stimulus-locked responses on human arm muscles reveal rapid neural pathway linking visual input to arm motor output. *European Journal of Neuroscience*, 32(6):1049–1057.

[Pruszynski et al., 2009] Pruszynski, J. A., Kurtzer, I., Lillicrap, T. P., and Scott, S. H. (2009). Temporal Evolution of " Automatic Gain-Scaling ". *Journal of Neurophysiology*, 102(2):992–1003.

[Pruszynski et al., 2011] Pruszynski, J. A., Kurtzer, I., Nashed, J. Y., Omrani, M., Brouwer, B., and Scott, S. H. (2011). Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature*, 478(7369):387–390.

[Pruszynski et al., 2008] Pruszynski, J. A., Kurtzer, I., and Scott, S. H. (2008). Rapid Motor Responses Are Appropriately Tuned to the Metrics of a Visuospatial Task. *Journal of Neurophysiology*, 100(1):224–238.

[Pruszynski and Scott, 2012] Pruszynski, J. A. and Scott, S. H. (2012). Optimal feedback control and the long-latency stretch response. *Experimental Brain Research*, 218(3):341–359.

[Qian et al., 2013] Qian, N., Jiang, Y., Jiang, Z.-P., and Mazzoni, P. (2013). Movement Duration, Fitts's Law, and an Infinite-Horizon Optimal Feedback Control Model for Biological Motor Systems. *Neural Computation*, 25(3):697–724.

[Ratcliff, 1978] Ratcliff, R. (1978). A theory of memory retrival. *Psychological Review*, 85(2):59–108.

[Ratcliff et al., 2007] Ratcliff, R., Hasegawa, Y. K., Hasegawa, R. P., Smith, P. L., and Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, 97(2):1756–1774.

[Reddi et al., 2003] Reddi, B. A. J., Asrress, K. N., and Carpenter, R. H. S. (2003). Accuracy, information and reponse time in a saccadic decision task. *Journal of Neurophysiology*, 90:3538–3546.

[Reddi and Carpenter, 2000] Reddi, B. A. J. and Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, 3(8):827–830.

[Reed, 1973] Reed, A. V. (1973). Speed-accuracy tradeoff in memory recognition. *Science*, 181(4099):574–576.

[Reppert et al., 2015] Reppert, T. R., Lempert, K. M., Glimcher, P. W., and R., S. (2015). Modulation of saccade vigor during value-based decision making. *Journal of Neuroscience*, 35(46):15369–15378.

[Reppert et al., 2018] Reppert, T. R., Rigas, I., Herzfeld, D. J., Sedaghatnejad, E., Komogortsev, O., and Shadmehr, R. (2018). Movement vigor as a traitlike attribute of individuality. *Journal of Neurophysiology*, 120(2):741–757.

[Retschechtko and Pruszynski, 2020] Retschechtko, S. and Pruszynski, J. A. (2020). Voluntary modification of rapid tactile-motor responses during reaching differs from its visuomotor counterpart. *Journal of Neurophysiology*, 124(1):284–294.

[Rigoux and Guigon, 2012] Rigoux, L. and Guigon, E. (2012). A model of reward- and effort-based optimal decision making and motor control. *PLoS Computational Biology*, 8(10):e1002716.

[Rinberg et al., 2006] Rinberg, D., Koulakov, A., and Gelperin, A. (2006). Speed-accuracy tradeoff in olfaction. *Neuron*, 51:351–358.

[Roitman and Shadlen,] Roitman, J. D. and Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combine visual discrimination reaction time task. *Journal of Neuroscience*, 22(21):9475–9489.

[Roll et al., 1989] Roll, J. P., Vedel, J. P., and Ribot, E. (1989). Alteration of proprioceptive messages induced by tendon vibration in man: a microneurographic study. *Experimental Brain Research*, 79:213–222.

[Roll et al., 1986] Roll, R., Bard, C., and Paillard, J. (1986). Head orienting contributes to the directional accuracy of aiming at distant targets. *Human movement science*, 5:359–371.

[Rossetti et al., 1995] Rossetti, Y., Desmurget, M., and Prablanc, C. (1995). Vectorial coding of movement: vision, proprioception, or both? *Journal of Neurophysiology*, 74(1):457–463.

[Rossetti et al., 1994] Rossetti, Y., Tadary, B., and Prablanc, C. (1994). Optimal contributions of head and eye positions to spatial accuracy in man tested by visually directed pointing. *Experimental Brain Research*, 97:487–496.

[Rothwell et al., 1980] Rothwell, J., Traub, M., and Marsden, C. D. (1980). Influence of voluntary intent on the human long-latency stretch reflex. *Nature*, 286(1):496–498.

[Rushworth et al., 1997] Rushworth, M. F., Nixon, P. D., and Passingham, R. E. (1997). Parietal cortex and movement. i . movement selection and reaching. *Experimental Brain Research*, 117(2):292–310.

[Saal and Bensmaia, 2014] Saal, H. P. and Bensmaia, S. J. (2014). Touch is a team effort : interplay of submodalities in cutaneous sensibility. *Trends in Neuroscience*, 37(12):689–697.

[Sabes et al., 1998] Sabes, P. N., Jordan, M. I., and Wolpert, D. M. (1998). The Role of Inertial Sensitivity in Motor Planning. *The Journal of Neuroscience*, 18(15):5948–5957.

[Sabes and Jordan, 1997] Sabes, P. S. and Jordan, M. I. (1997). Obstacle avoidance and a perturbation sensitivity model for motor planning. *Journal of Neuroscience*, 17(18):7119–7128.

[Sackaloo et al., 2014] Sackaloo, K., Otr, L., Strouse, E., Otr, L., Rice, M. S., and Otr, L. (2014). Degree of Preference and Its Influence on Motor Control When Reaching for Most Preferred , Neutrally Preferred , and Least Preferred Candy. *OTJR: Occupation, Participation and Health*, 35(2):81–88.

[Sainburg et al., 2003] Sainburg, R. L., Lateiner, J. E., Latash, M. L., and Bagesteiro, L. B. (2003). Effects of altering initial position on movement direction and extent. *Journal of Neurophysiology*, 89(1):401–415.

[Sarlegna et al., 2003] Sarlegna, F., Blouin, J., Bresciani, J. P., Bourdin, C., Vercher, J. L., and Gauthier, G. M. (2003). Target and hand position information in the online control of goal-directed arm movements. *Experimental Brain Research*, 151(4):524–535.

[Sarlegna and Mutha, 2014] Sarlegna, F. R. and Mutha, P. K. (2014). The influence of visual target information on the online control of movements. *Vision Research*.

[Saunders and Knill, 2004] Saunders, J. A. and Knill, D. C. (2004). Visual Feedback Control of Hand Movements. *Journal of Neuroscience*, 24(13):3223–3234.

[Schaffelhofer et al., 2015] Schaffelhofer, S., Sartori, M., Scherberger, H., and Farina, D. (2015). Musculoskeletal representation of a large repertoire of hand grasping actions in primates. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(2):210–220.

[Scholz et al., 2000] Scholz, J. P., Schoner, G., and Latash, M. L. (2000). Identifying the control structure of multijoint coordination during pistol shooting. *Experimental Brain Research*, 135(3):382–404.

[Schulman et al., 2016] Schulman, J., Moritz, P., Levine, S., Jordan, M. I., and Abbeel, P. (2016). High-dimensional continuous control using generalized advantage estimation. *arXiv*.

[Schultz, 2006] Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57:87–115.

[Schultz et al., 1993] Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, 13(3).

[Schultz et al., 1997] Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

[Scott, 2004] Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7):532–546. ISBN: 1471-0048.

[Scott, 2012] Scott, S. H. (2012). The computational and neural basis of voluntary motor control and planning. *Trends in Cognitive Sciences*, 16(11):541–549.

[Scott, 2016] Scott, S. H. (2016). A functional taxonomy of bottom-up sensory feedback processing for motor actions. *Trends in Neurosciences*, 39(8):512–526.

[Selen et al., 2012] Selen, L. P. J., Shadlen, M. N., and M, W. D. (2012). Deliberation in the motor system : Reflex gains track evolving evidence leading to a decision. *Journal of Neuroscience*, 32(7):2276–2286.

[Shadmehr et al., 2016] Shadmehr, R., Huang, H. J., and Ahmed, A. A. (2016). A representation of effort in decision-making and motor control. *Experimental Brain Research*, 26(14):1929–1934.

[Shadmehr et al., 2010a] Shadmehr, R., Jean-Jacques, O. d. X., Xu-Wilson, M., and Shih, T.-y. (2010a). Temporal Discounting of Reward and the Cost of Time in Motor Control. *Journal of Neuroscience*, 30(31):10507–10516.

[Shadmehr and Krakauer, 2008] Shadmehr, R. and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185(3):359–381.

[Shadmehr and Mussa-Ivaldi, 1994] Shadmehr, R. and Mussa-Ivaldi, F. a. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5):3208–3224.

[Shadmehr et al., 2019] Shadmehr, R., Reppert, T. R., Summerside, E. M., Yoon, T., and Ahmed, A. A. (2019). Movement vigor as a reflection of subjective economic utility. *Trends in Neurosciences*, 42(5):323–336.

[Shadmehr et al., 2010b] Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010b). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33(1):89–108.

[Shemmell et al., 2009] Shemmell, J., Je, H. A., and Perreault, E. J. (2009). The differential role of motor cortex in stretch reflex modulation induced by changes in environmental mechanics and verbal instruction. *Journal of Neuroscience*, 29(42):13255–13263.

[Shenoy et al., 2013] Shenoy, K. V., Sahani, M., and Churchland, M. M. (2013). Cortical Control of Arm Movements: A Dynamical Systems Perspective. *Annual Review of Neuroscience*, 36(1):337–359.

[Singh and Scott, 2003] Singh, K. and Scott, S. H. (2003). A motor learning strategy reflects neural circuitry for limb control. *Nature Neuroscience*, 6:399–403.

[Smeets and Brenner, 2003] Smeets, J. and Brenner, E. (2003). Fast corrections of movements with a computer mouse. *Spatial Vision*, 16(3):365–376.

[Smith et al., 2006] Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLOS Biology*, 4(6):1035–1043.

[Sober and Sabes, 2003] Sober, S. J. and Sabes, P. N. (2003). Multisensory integration during motor planning. *Journal of Neuroscience*, 23(18):6982–6992.

[Soechting, 1984] Soechting, J. F. (1984). Effect of target size on spatial and temporal characteristics of a pointing movement in man. *Experimental Brain Research*, 54(1):121–132.

[Soechting and Lacquaniti, 1981] Soechting, J. F. and Lacquaniti, F. (1981). Invariant characteristics of a pointing arm movement in man. *Journal of Neuroscience*, 1(7):710–720.

[Soechting and Lacquaniti, 1983] Soechting, J. F. and 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.

[Spence et al., 2000] Spence, C., Pavani, F., and Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology*, 26(4):1298–1319.

[Stein et al., 1995] Stein, R. B., Hunter, I. W., Lafontaine, S. R., and Jones, L. A. (1995). Analysis of short-latency reflexes in human elbow flexor muscles. *Journal of Neurophysiology*, 73:1900–1911.

[Striemer et al., 2010] Striemer, C. L., Yukovsky, J., and Goodale, M. A. (2010). Can intention override the "automatic pilot"? *Experimental Brain Research*, 202:623–632.

[Summerside et al., 2018] Summerside, E. M., Shadmehr, R., and Ahmed, A. A. (2018). Control of Movement Vigor of reaching movements : reward discounts the cost of effort. *Journal of Neurophysiology*, 119(6):2347–2357.

[Sutton, 2000] Sutton, R. S. (2000). Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063.

[Sutton and Barto, 2018] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction, 2nd edition*. MIT Press.

[Tachibana and Hikosaka, 2012] Tachibana, Y. and Hikosaka, O. (2012). The primate ventral pallidum encodes expected reward value and regulates motor action. *Neuron*, 76(4):826–837.

[Therrien et al., 2016] Therrien, A. S., Wolpert, D. M., and Bastian, A. J. (2016). Effective Reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, 139(1):101–114.

[Thompson et al., 2009] Thompson, A. K., Chen, X. Y., and Wolpaw, J. R. (2009). Acquisition of a simple motor skill: task-dependent adaptation plus long-term change in the human soleus h-reflex. *Journal of Neuroscience*, 29(18):5784–8792.

[Thura et al., 2012] Thura, D., Beauregard-Racine, J., Fradet, C.-W., and Cisek, P. (2012). Decision making by urgency-gating: Theory and experimental support. *Journal of Neurophysiology*, 108(11):2912–2930.

[Thura and Cisek, 2014] Thura, D. and Cisek, P. (2014). Deliberation and commitment in the premotor cortex and primary motor cortex during dynamic decision making. *Neuron*, 81(6):1401–1416.

[Thura and Cisek, 2017] Thura, D. and Cisek, P. (2017). The basal ganglia do not select reach targets but control the rugency of commitment. *Neuron*, 95(5):991–993.

[Todorov, 2004] Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7:907–915.

[Todorov, 2005] 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 and Jordan, 2002] Todorov, E. and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235.

[Todorov and Li, 2005] Todorov, E. and Li, W. (2005). A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 300–306 vol. 1.

[Togo et al., 2017] Togo, S., Yoshioka, T., and Imamizu, H. (2017). Control strategy of hand movement depends on target redundancy. *Scientific Reports*, 7:1–7.

[Trommershäuser et al., 2005] Trommershäuser, J., Gepshtein, S., Maloney, L. T., Landy, M. S., and Banks, M. S. (2005). Optimal compensation for changes in task-relevant movement variability. *Journal of Neuroscience*, 25(31):7169–7178.

[Trommershäuser et al., 2003] Trommershäuser, J., Maloney, L. T., and Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision*, 16(3-4):255–275.

[Trommershäuser et al., 2008] Trommershäuser, J., Maloney, L. T., and Landy, M. S. (2008). Decision making , movement planning and statistical decision theory. *Trends in Cognitive Sciences*, 12(8):291–297.

[Turner and Desmurget, 2010] Turner, R. S. and Desmurget, M. (2010). Basal ganglia contributions to motor control: A vigorous tutor. *Current Opinion in Neurobiology*, 70(6):704–716.

[Uno et al., 1989] Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61:89–101.

[van Beers et al., 2002] van Beers, R., Baraduc, P., and Wolpert, D. M. (2002). Role of uncertainty in motor control. *Philosophical transactions of the Royal Society B*, 357:1137–1145.

[Vanden Abeele et al., 1993] Vanden Abeele, S., Delreux, V., Crommelinck, M., and Roucoux, A. (1993). Role of eye and hand initial position in the directional coding of reaching. *Journal of Motor Behavior*, 25(4):280–287.

[Verrier, 1985] Verrier, M. C. (1985). Alterations in H reflex magnitude by variations in baseline EMG excitability. *Electroencephalogram Clin Neurophysiology*, 60:492–499.

[Vindras et al., 2005] Vindras, P., Desmurget, M., and Viviani, P. (2005). Error parsing in visuomotor pointing reveals independent processing of amplitude and direction. *Journal of Neurophysiology*, 94(2):1212–1224.

[Wald, 1945] Wald, A. (1945). Sequential tests of statistical hypothesis. *The Annals of Mathematical Statistics*, 16(2):117–186.

[Wald and Wolfowitz, 1948] Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339.

[Wang et al., 2008] Wang, Q.-G., Ye, Z., Cai, W.-J., and Hang, C.-C. (2008). *PID Control for multivariable processes*. Springer.

[Wang et al., 1997] Wang, Q.-G., Zou, B., Lee, T.-H., and Bi, Q. (1997). Auto-tuning of multivariable pid controllers from decentralized relay feedback. *Automatica*, 33(3):319–330.

[Wächter et al., 2009] Wächter, T., Lungu, O. V., Liu, T., Willingham, D. T., and Ashe, J. (2009). Differential effect of reward and punishment on procedural learning. *Journal of Neuroscience*, 29(2):436–443.

[Weber et al., 2013] Weber, A. L., Saal, H. P., Lieber, J. D., Cheng, J.-W., Manfredi, L. R., Dammann, J. F., and Bensmaia, S. J. (2013). Spatial and temporal codes mediate the tactile perception of natural textures. *PNAS*, 110(42):17107–17112.

[Weiler et al., 2015] Weiler, J., Gribble, P. L., and Pruszynski, J. A. (2015). Goal-dependent modulation of the long-latency stretch response at the shoulder, elbow and wrist. *Journal of Neurophysiology*, 114(6):3242–3254.

[Weiler et al., 2019] Weiler, J., L, G. P., and Pruszynski, J. A. (2019). Spinal stretch reflexes support efficient hand control. *Nature Neuroscience*, 22:529–533.

[Weiler et al., 2016] Weiler, J., Saravanamuttu, J., Gribble, P. L., and Pruszynski, J. A. (2016). Coordinating long-latency stretch responses across the shoulder, elbow and wrist during goal-directed reaching. *Journal of Neurophysiology*, 116(5):2236–2249.

[Westheimer, 1984] Westheimer, G. (1984). Spatial vision. *Annual Review of Psychology*, 35:201–226.

[Wickelgren, 1977] Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41:67–85.

[Wolpaw, 1983] Wolpaw, J. R. (1983). Adaptive plasticity in the primate spinal stretch reflex: reversal and re-development. *Brain Research*, 278(1-2):299–304.

[Wolpert et al., 1995] Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). Are arm trajectories planned in kinematics or dynamic coordinates ? an adaptation study. *Experimental Brain Research*, 103:460–470.

[Wolpert et al., 1998] Wolpert, D. M., Goodboy, S. J., and Husain, M. (1998). Maintanining internal representations: the role of the human superior parietal lobe. *Nature Neuroscience*, 1(6):529–533.

[Wolpert and Landy, 2012] Wolpert, D. M. and Landy, M. S. (2012). Motor control is decision-making. *Current opinion in neurobiology*, 22(6):996–1003.

[Wong et al., 2017] Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., and Krakauer, J. W. (2017). Reaction times can reflect habits rather than computations. *eLife*, 6:1–18.

[Wong et al., 2015] Wong, A. L., Haith, A. M., and Krakauer, J. W. (2015). Motor planning. *Neuroscientist*, 21(4):385–398.

[Wong et al., 2021] Wong, J. D., Cluff, T., and Kuo, A. D. (2021). The energetic basis for smooth human arm movements. *bioRxiv*.

[Woodman et al., 2008] Woodman, G. F., Kang, M.-S., Thomson, K., and Schall, J. D. (2008). The effect of visual search efficiency on response preparation. *Psychological Science*, 19(2):128–136.

[Woodworth, 1889] Woodworth, R. S. (1889). Accuracy of voluntary movement. *The psychological Review: Monograph Supplements*, 3(3).

[Xiao, 2017] Xiao, Y. (2017). A fast algorithm for two-dimensional Kolmogorov-Smirnov two sample tests. *Computational Statistics and Data Analysis*, 105:53–58.

[Xu-Wilson et al., 2009] Xu-Wilson, M., Zee, D. S., and Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, 196(4):475–481.

[Yang et al., 2011] Yang, L., Michaels, J. A., Pruszynski, J. A., and Scott, S. H. (2011). Rapid motor responses quickly integrate visuospatial task constraints. *Experimental Brain Research*, 211(2):231–242.

[Yoon et al., 2018] Yoon, T., Geary, R. B., Ahmed, A. A., and Shadmehr, R. (2018). Control of movement vigor and decision making during foraging. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44):10476–10485.

[Zhariev and MacKenzie, 2007] Zhariev, M. A. and MacKenzie, C. L. (2007). Grasping at 'thin air': multimodal contact cues for reaching and grasping. *Experimental Brain Research*, 180:69–84.

[Ziegler and Nichols, 1942] Ziegler, J. G. and Nichols, N. (1942). Optimum settings for automatic controllers. *Transactions ASME*, 64(11).