A model of reward- and effort-based optimal decision making and motor control

Abbreviated title: Optimal decision and action

Lionel Rigoux[1,2], Emmanuel Guigon[1,2]
[1] UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France
[2] CNRS, UMR 7222, ISIR, F-75005, Paris, France

Correspondence to:
Emmanuel Guigon
Institut des Systèmes Intelligents et de Robotique
UPMC — CNRS / UMR 7222
Pyramide Tour 55 - Boîte Courier 173
4 Place Jussieu
75252 Paris Cedex 05, France
Fax: 33 1 44 27 63 82
Tel: 33 1 44 27 51 45
email: emmanuel.guigon@gmail.com

Abstract

Costs (e.g. energetic expenditure) and benefits (e.g. food) are central determinants of behavior. In ecology and economics, they are combined to form a utility function which is maximized to guide choices. This principle is widely used in neuroscience as a normative model of decision and action, but current versions of this model fail to consider how decisions are actually converted into actions (i.e. the formation of trajectories). Here, we describe an approach where decision making and motor control are optimal, iterative processes derived from the maximization of the discounted, weighted difference between expected rewards and foreseeable motor efforts. The model accounts for decision making in cost/benefit situations, and detailed characteristics of control and goal tracking in realistic motor tasks. As a normative construction, the model is relevant to address the neural bases and pathological aspects of decision making and motor control.

Author summary

Behavior is made of decisions and actions. The decisions are based on the costs and benefits of potentials actions, and the chosen actions are executed through the proper control of body segments. The corresponding processes are generally considered in separate theories of decision making and motor control, which cannot explain how the actual costs and benefits of a chosen action can be consistent with the expected costs and benefits involved at the decision stage. Here, we propose an overarching optimal model of decision and motor control based on the maximization of a mixed function of costs and benefits. The model provides a unified account of decision in cost/benefit situations (e.g. choice between small reward/low effort and large reward/high effort options), and motor control in realistic motor tasks. The model appears suitable to advance our understanding of the neural bases and pathological aspects of decision making and motor control.

Introduction

Consider a simple living creature that needs to move in its environment to collect food for survival (foraging problem; [1]). For instance, it can have to choose between a small amount of food at a short distance and a larger amount at a longer distance [2, 3]. These two choices should not in general be equivalent as they differ by the proposed benefit (amount of food), the cost of time (temporal discounting of the benefit), and the cost of movement (energetic expenditure) [4-6]. To behave appropriately in its environment, our creature should be able to: 1. make decisions based on the estimated costs and benefits of actions; 2. translate selected actions into actual movements in a way which is consistent with the decision process, i.e. the criterion used *a priori* for decision should be backed up *a posteriori* by the measured costs and benefits of the selected action; 3. update its behavior at any time during the course of action as required by changes in the environment (e.g. removal or change in the position of food).

Most theories of decision making and motor control do not account for these characteristics of behavior. The main reason for this is that decision and control are essentially blind to each other in the proposed frameworks [7]. On the one hand, standard theories of decision making [8] rely on value-based processes (e.g. maximization of expected benefit), and fail to integrate the cost of physical actions into decisions [9]. On the other hand, modern theories of motor control are cast in the framework of optimal control theory, and propose to elaborate motor commands using a cost-based process (e.g. minimization of effort), irrespective of the value of actions [10, 11]. An interesting exception is the model proposed by Trommershäuser et al. [12-14] which casts into a Bayesian framework the observation that at least one aspect of motor

control (intrinsic motor variability) is optimally integrated into decision making processes.

Here, we consider a normative approach to decision making and motor control derived from the theory of *reinforcement learning* (RL; [15-17]), i.e. goals are defined by spatially located time-discounted rewards, and decision making and motor control are optimal processes based on the maximization of utility, defined as the discounted difference between benefits (reward) and costs (of motor commands). The proposed mechanism concurrently provides a criterion for choice among multiple actions, and an optimal control policy for execution of the chosen action. We show that: 1. The model accounts for decision making in cost/benefit situations, and characteristics of control in realistic motor tasks; 2. Parameters that govern the model can explain the perviousness of these behaviors to motivational and task-related influences (precision, instructions, urgency). As a normative construction, the model can be considered as a prescription of what the nervous system should do [18], and is thus relevant to address and discuss the neural bases and pathological aspects of decision making and motor control. In particular, we focus on the role of dopamine (DA) whose implication in decision making, motor control and reward/effort processing has been repeatedly emphasized [2, 6, 19-22].

Results

The proposed model is a model for decision and action. It is based on an objective function representing a trade-off between expected benefits and foreseeable costs of potential actions (Fig. 1A and Eq. 4; see **Materials and Methods**). Maximization of this function attributes a utility to each action, which can be used for a decision process,

and generate a control policy to carry out the action (Eq. 6). Our goal is two-fold. First, we show that the model accounts for decision making in cost/benefit situations, and control in realistic motor tasks. Second, we show that the model makes sense from a psychological and neural standpoint. As a preliminary, we describe parameters that are central to the functioning of the model.

Nature of the parameters

The model contains five parameters ($\mathbf{x}^*$, r, ρ, ε, γ; Eqs. 5 and 6). Parameter $\mathbf{x}^*$ specifies the location of the goal to be pursued, and acts as a classic boundary condition for a control policy. Parameter r is a value attached to the goal that can correspond to a reward on an objective scale (e.g. amount of food, amount of money), or to any factor that modulates the pursuit and achievement of goals (e.g. interest, attractiveness, difficulty, ...). For pure motor tasks in which there is no explicit reward, we will assume that r corresponds to one of these factors (see **Discussion**). $\mathbf{x}^*$ and r are parameters related to the specification of a task, and will be called *task* parameters.

For the purpose of decision and action, a reward value needs to be translated into an internal currency which measures "how much a reward is rewarding" (parameter ρ). A subject may not attribute the same value to food if he is hungry or satiated, and the same value to money if he plays Monopoly or trades at the stock exchange. r and ρ are redundant in the sense that only their product matters (Eq. 6), but we keep both of them because their meaning is different.

Parameter ε is a scaling factor that expresses "how much an effort is effortful". A subject may not attribute the same value to effort if he is rested or exhausted. ρ and ε are redundant in the sense that only their ratio matters (Eq. 6), but we keep both of them because their meaning is different, and they can be regulated differently (e.g. level of

wealth vs level of fatigue). In general, we consider variations in the ratio ρ/ε, that we call *vigor* factor in the following.

Parameter γ is a discount factor on reward and effort. It is both a computational parameter that is necessary to the formulation of the model, and a factor related to the process by which delayed or far away reinforcers lose value [3, 23]. Note that a decrease in γ corresponds to faster discount.

In the following, ρ, ε, and γ are called *internal* parameters, to indicate that they are not directly specified by the external environment, but correspond to a subjective valuation of concrete influences in the body and the environment. These parameters are allowed to vary to explore their role in the model. To provide a neural interpretation of the model, we tentatively relate effects of these variations to identified physiological elements.

We note that the principle of the model is independent of the values of the parameters, i.e. the decision process and the control policy are generic characteristics of the model.

Decision making in a cost/benefit situation

The model provides a normative criterion for decision making when choices involve different costs and benefits. To explore this issue, we considered the simple situation depicted in Fig. 2A: a small reward at a short distance (reference distance) and a larger reward at a variable distance (test distance). Distance is used here as a way to modulate the required effort level. Simulations were run with Object I in the absence of noise. As the test distance increased, the effort to obtain the larger reward increased, and the utility decreased (Fig. 2B). Beyond a given distance (*indifference point*), the utility became smaller than the reference utility. Thus the indifference point separated two

regions corresponding to a preference for the large reward/high effort and the small reward/low effort. This result corresponds to a classic observation in cost/benefit choice tasks [4, 6].

The model further states that the same parameters underlie both decision and movement production. To test this idea, we modeled the experiment reported by Stevens et al. [3] [referred as Stevens in the following], in which the behavior of two species of monkey (marmoset and tamarin) was assessed in the choice situation of Fig. 2A. The monkeys had to choose between one reward at 35 cm, and three rewards at 35-245 cm (distances 1 to 7). Stevens reported the choice behavior of the monkeys (Fig. 2 in Stevens) as well as the durations of chosen actions (Fig. 3 in Stevens). The modeling principle is the following. We consider that the behavior of a monkey is determined by two parameters: a vigor factor (ρ/ε) and a discount factor (γ). The question is: if we infer these parameters from the displacement duration of the monkey, can we explain its choice behavior? An important issue is the underlying determinant of amplitude/duration data (Fig. 3 in Stevens). There is strong experimental evidence for the existence of a linear relationship between distance and duration for locomotor displacements ([24-27]; see also [28] with fish). This observation suggests that two parameters could be sufficient to capture covariations between displacement amplitudes and durations.

For Object I, we have an analytic formula for optimal movement duration $T^*(A,r,\rho/\varepsilon,\gamma)$ as a function of movement amplitude (A), reward (r), vigor (ρ/ε) and discount (γ) (see **Materials and Methods**). From Fig. 3 in Stevens, we also obtained the duration of displacement T (mean±s.e.m of the individual mean performances across the population) for each species in two conditions: one reward ($r_1 = 1$) located at

A_1 = 0.35 m (marmoset: T_1 = .75±.061 s, tamarin: T_1 = .66±.047 s), and three rewards (r_2 = 3) at A_2 = 2.45 m (marmoset: T_2 = 1.84±.082 s, tamarin: T_2 = 1.32±.050 s).

We randomly drew pairs of movement duration (one for each condition) from a Gaussian distribution specified by the mean and sd (= s.e.m x sqrt(N), N = 4) given above, thus generating for each species a set of synthetic monkeys (n = 100). For each sample monkey, we obtained a unique value of vigor and discount factors [two unknowns: ρ/ε and γ; two equations: $T_1 = T^*(A_1,r_1,\rho/\varepsilon,\gamma)$ and $T_2 = T^*(A_2,r_2,\rho/\varepsilon,\gamma)$]. The corresponding parameters are shown in Fig. 2C. The two synthetic species were clearly associated with distinct regions of the parameter space, the marmosets being more sensitive to effort than the tamarins. It should be noted that Fig. 2C does not mean that there exists a redundancy between the two parameters: in fact, each point of the clouds corresponds to a different displacement behavior, i.e. different distance/duration relationships. The correlation between the parameters suggests a potential lack of specificity of the duration measurements for our method to parsimoniously characterize the populations. However, although it would be possible to tighten our predictions with more structured data (e.g. estimated parameters based on individual behavior), it is unnecessary to reveal a clear cut dissociation between the two species.

Then we computed for each monkey (i.e. for each set of parameters shown in Fig. 2C) the utility of the different options (1 reward/35 cm, 3 rewards/35-245 cm). The two sets of parameters produced different indifference points (Fig. 2D). Specifically, the majority of marmosets, in contrast with tamarins, showed an inversion in their preferences within the tested range of distances (< 2.45 m).

To determine the choice behavior of the monkeys from option utilities, we calculated the probability to choose the large reward at the different distances vs the small reward at the shortest distance using a softmax rule

$$P(\text{large}) = \exp(J_{\infty}^{\text{large}}/\beta)/[\exp(J_{\infty}^{\text{large}}/\beta) + \exp(J_{\infty}^{\text{small}}/\beta)],$$

where $J_{\infty}^{\text{large}}$ and $J_{\infty}^{\text{small}}$ are the utilities for the large reward and small reward options, respectively, and β a temperature parameter which represents the degree of randomness of the action selection. It should be noted that the softmax transform is not a part of the model, but a way to translate utilities into choice proportions, using the natural principle that different option utilities should lead to a proportion near 1 (or 0), and equal option utilities to a proportion of 0.5. The parameter β, which had no qualitative effect on the predicted preferences, was selected for each monkey to fit the data from Stevens. The model quantitatively reproduced the empirical results in the decision task for the two monkey species (Fig. 2E). Some outliers exhibited a less characteristic behavior (whiskers in Fig. 2D) due to some imprecisions in our estimation. However, these marginal profiles were very scarce, and did not undermine our general results (see confidence interval; Fig. 2E).

To assess more precisely the ability of the model to predict the choices, we performed a detailed analysis over the two sets of simulated utilities (not over choices, to rule out any confound induced by β). We found that distance to the large reward modulated the utility of the large reward for both species, and that: 1. for tamarins, the large reward option had a larger utility than the small reward option for all distances; 2. for marmosets, the large reward option had a larger utility than the small reward option only for test distances strictly smaller than 210 cm. These results exactly parallel the effects found by Stevens, and show that the model can quantitatively predict the

inversion of preferences of the different species. This further supports the hypothesis that the same process governs decision making and action in a cost/benefit choice situation.

Control in realistic motor tasks

The model reproduced basic characteristics of motor behavior, as expected from the close relationship with previous optimal control models [10, 11, 29, 30]. Simulations were run with Object IIIa (two-joint planar arm) in the absence of noise. The internal parameters (ρ/ε and γ) were chosen to obtain a range of velocities compatible with observations on arm movements, and were kept constant for simulations of motor control task (Figs. 3, 4, 5). Their values had no qualitative influence on the reported results. Movements of different amplitudes (Fig. 3A) and in different directions (Fig. 3B) were considered. Simulated trajectories were straight (Fig. 3A,B) with a bell-shaped velocity profile (Fig. 3C, inset). Movement duration emerged implicitly corresponding to the best compromise between discounted rewards and efforts. Accordingly, duration was a function of movement amplitude (amplitude/duration scaling law; Fig. 3C), and movement direction (Fig. 3D, *plain line*). In fact, the influence of direction was related the inertial anisotropy of the arm (Fig. 3D, *dotted line*). Scaling was also observed for peak velocity and peak acceleration (not shown). These results are consistent with experimental observations [31].

Unexpected events can perturb an ongoing action, and prevent a planned movement to reach its goal. Typical examples are sudden changes in target location [29] or mechanical alteration of limbs dynamics [32]. In these experiments, participants correct their movements and proceed to the goal by smoothly modifying the kinematics of their arm and the duration of the action. In the model, movement duration is not fully

specified in advance, but emerges from an online feedback process concerned only by the remaining effort necessary to get a reward. We wanted to test if this property could explain motor control when movement execution requires flexibility to deal with unforeseen perturbations.

In the experiment of Liu and Todorov [29], the target location jumped unpredictably during the reach. This caused a lengthening of movement duration which increased with the time elapsed between movement onset and perturbation onset (perturbation time; Fig. 1g in [29]), and systematic modifications of trajectory (Fig. 1a in [29]) and velocity profile (Fig. 1b in [29]). We simulated this task with Object IIIa by changing the goal position ($\mathbf{x}^*$) in the controller at different times (perturbation time+Δ, to account for delayed perception of the change). The parameters of the model were estimated from unperturbed trials. The model quantitatively reproduced trajectory formation (Fig. 4A; Fig. 1a in [29]), velocity profiles (Fig. 4B; Fig. 1b in [29]), and the effect of perturbation time on movement duration (Fig. 4C; Fig. 1g in [29]). Liu and Todorov [29] have proposed an optimal feedback control model to explain their results. However, in their approach, the duration of perturbed movements was not an emergent property of the model, and they used experimentally measured durations in their simulations. Later in their article, they described a different model, including a cost of time, which was potentially able to predict the duration of perturbed movements, but this model was not used to explain their initial target jump data.

In the experiment of Shadmehr and Mussa-Ivaldi [32], participants performed reaching movements using a robotic device that exerted a force on their arm, i.e. altered the dynamic of their limb and continuously deflected the arm from its intended trajectory. Initial exposure to the perturbation induced deviations from straight line

trajectories with typical hook-like final corrections (Fig. 7 in [32]), and multiple peak velocity profiles (Fig. 10 in [32]). We simulated this task with Object IIIb in the presence of a velocity-dependent force field. The controller was unaware of the presence of the force field. The parameters were those used in the preceding simulations (Figs. 3 and 4), and were appropriate to fit unperturbed trials. Unperturbed velocity profiles are shown for 4 directions in Fig. 5A. From the interplay between the naïve controller and the altered arm dynamics emerged curved trajectories with typical hooks (Fig. 5B), and multi-peaked velocity profiles (Fig. 5C), which are qualitatively similar to the experimental data.

These results illustrate how a unique set of parameters, and thus a unique controller, explains both normal trajectory formation, and complex updating of motor commands and trajectories when participants face unexpected perturbations. The same mechanisms (optimality, feedback control, implicit determination of duration) underlie basic motor characteristics (scaling law), and flexible control and goal tracking in complex situations.

Modulation of decision making and motor control

The model is governed by the vigor (ρ/ε) and discount (γ) factors that can modulate both the decision process and the control policy (Eq. 6).

Decision making in a cost/benefit situation (Fig. 2A) was characterized by a threshold that delineates choice preference between small reward/low effort and large reward/high effort options (Fig. 2B). We observed a shift of the decision criterion toward the small reward/low effort option for a decreased vigor (lower ρ/ε; Fig. 6A), or a steepened discount (lower γ; Fig. 6B). Interestingly, the shift was accompanied by a decreased velocity in the former case (Fig. 6C), and an increased velocity in the latter

(Fig. 6D). Note that the parameters were different from those used in Fig. 2, and were chosen here to obtain a range of velocities compatible with observations on arm movements. This choice had no influence on the results. This result is especially interesting since it reveals a dissociation between the influence of vigor and discount on decision making and motor control. The effects of vigor, but not discount, resemble the shift of decision criterion toward small reward/low effort options [2, 6, 20], and the decrease in velocity [2] observed in rat's behavior following systemic injection of dopamine receptor antagonists or DA depletion in the ventral striatum.

Motor control was characterized by scaling laws (Fig. 3C). Each factor, by its variation, defined a family of amplitude/duration scaling laws. For instance, a decrease in vigor induced an upward shift of the scaling law (Fig. 6C). Consistent with the influence of vigor described above, this result could correspond to the widely reported preservation and shift of amplitude/duration (and amplitude/velocity) scaling laws across DA manipulations and basal ganglia lesions in animals [33-36], and basal ganglia disorders in humans (bradykinesia; [37-39]). However, this interpretation is tentative as the shifts induced by vigor and discount were qualitatively similar (Fig. 6C,D; see **Discussion**).

Along the scaling laws defined by each factor (Fig. 6C,D), amplitude, duration and variability varied in a concerted way that conformed to Fitts' law [40, 41], i.e. movement duration is a function of the index of difficulty (i.e. $\log_2(2A/W)$, where A is the amplitude and W the endpoint variability; Fig. 7A). We note that the underlying pattern of spatiotemporal variability had two peaks, one around peak velocity and the other near the end of the movement (Fig. 7B), and is consistent with experimental observations (although the temporal profiles are usually cut before variability starts to

return toward premovement levels; [29, 42, 43]). These results show that the vigor and discount factors can induce modulations of movement duration and scaling laws that might correspond to experimentally identified elements (see above and **Discussion**) while strictly obeying to a robust and ubiquitous law of motor control. Interestingly, for a given amplitude, any of these factors can act as an internal representation of a target size (Fig. 7C), i.e. it specifies a control policy that can instantaneously elaborate a movement of a given precision. It should be noted that there exist numerous models of Fitts' law in the literature [30, 44-46]. Our purpose here is not to propose a new model, but simply to check that Fitts' law can properly emerge from the proposed framework.

Overall, these results show that the internal parameters modulate decision making and motor control in a way that makes sense from a physiological and psychological point of view.

Discussion

We have presented a computational framework that describes decision making and motor control as an ecological problem. The problem was cast in the framework of reinforcement learning, and the solution formulated as an optimal decision process and an optimal control policy. The resulting model successfully addressed decision making in cost/benefit situations and control in realistic motor tasks.

Disclaimer

The proposed model is not intended to be a general theory of decision making and motor control, which may not be feasible (e.g. [47]), but a more modest theory for cost/benefit situations, i.e. specific situations in which expected benefits and foreseeable physical costs of potential actions have to be evaluated and balanced. Accordingly, the

model is not concerned with classic issues of risk and uncertainty which have been thoroughly addressed in studies of Trommershäuser and colleagues [12-14].

Previous models

Our model is closely related to previous works in the field of decision making and motor control. The central idea derives from optimal feedback control theory [10], and continuous time reinforcement learning [16, 17]. Several modeling studies have proposed modified versions of the optimal control approach to explain movement duration and amplitude/duration scaling laws [29, 48-50]. The common idea is to consider a compromise between a cost of time (which increases with movement duration), and a cost of action (which decreases with movement duration; [29, 48-50]). In a different framework, Niv et al. [51] proposed a compromise between a "cost of acting quickly" and a cost of "getting the reward belatedly". In these studies, the two costs varied in opposite directions with time, and their sum had a minimum value corresponding to an optimal behavioral timing (movement duration or latency; e.g. Fig. 1B in [49]). Our model exploits the same formal idea (our Fig. 1A), but with two differences. First, the cost of time in the previous studies were chosen for specific, task-related purposes (e.g. minimize the loss of vision from image motion during a saccade in [49]; minimize the time it takes to get a target on the fovea with a saccade in [50]; see below for a further discussion on the cost of time). In our model, the cost of time derives from a general normative criterion. Second, optimization in the previous models involved only cost terms. In these approaches (e.g. [50]), a larger reward leads to a larger cost of time, thus producing a faster movements but also a lower utility, which is problematic if one wants to account for rational choices between actions. Indeed, none of these formalisms proposed to formulate motor control as a decision making problem.

In our model, the reward modulates a benefit term, i.e. a larger reward leads to a larger benefit. This latter approach may be more appropriate to address cost/benefit situations in behavioral studies [52, 53], and the differential sensitivity of costs and benefits to pharmacological manipulations [52].

A series of study by Trommershäuser and colleagues [12-14] has explored the connection between decision making and motor control. These studies showed that human participants make optimal motor decisions (where to point in a spatial reward/penalty landscape) that take into account their intrinsic motor variability. The results suggest that at least one aspect of motor control (variability) is integrated into decision making processes (see also [54]). Our study explores a different aspect of the interaction between decision making and motor control: the influence of motor costs. In the early publications of Trommershäuser and colleagues [12, 13], a biomechanical cost was introduced, but was not actually used as it was assumed to be constant. The model described in [12, 13] is a model of decision making, which solves a spatial gain/loss trade-off at a motor planning level, but not a model of motor control as it does neither explain how movements are actually produced following a decision, nor how motor variability is estimated for a use in the decision process. Our model is primarily a model of motor control, which solves a temporal reward/effort trade-off at a motor control level, but disregards the issue of uncertainty. In this sense, our approach and that developed by Trommershäuser and colleagues [55] are complementary, and both useful to disclose the relationships between decision making and motor control.

A central and novel aspect of the model is the integration of motor control into the decision process. This idea was not exploited in previous models because movement duration was fixed [13, 56]. Our model is close to the model proposed by Dean et al.

[57] (see below), as both models involve a trade-off between a time-decaying (reward) quantity and a time-increasing (accuracy in [51], minus effort in our model) quantity. However, the time-increasing quantity in [57] is derived from experimental data, and is not generated by the model, i.e. there is no normative account of the speed/accuracy relationship.

The model was described here in its simplest form. In particular, decision making was considered as a deterministic process. The scope of the model could easily be extended to address stochastic paradigms as in previous models [13, 56]. Utility needs to be replaced by mean (expected) utility or possibly mean-variance combinations [7]. Further extensions could involve subjective utilities. In fact, none of these modifications would alter the very principle of the model.

Decision making

An analysis of behavior in terms of costs and benefits has long since been usual in behavioral ecology [1], but has only recently been exploited in the study of choice behavior in the field of neuroscience [5, 52, 58]. There is now strong evidence that not only payoff but also cost in terms of time and physical effort are integrated in the valuation of actions during a decision process [2, 6, 52, 59]. The model captures this view using an objective function in which a temporal cost is represented by a discount factor on the payoff (reward), and an effort cost by the integrated size of motor commands. The strength of this function is that it is not merely an aggregation of cost and benefit terms [50], but it has a true normative and sequential dimension [16, 17] which gives a consistent account of decision making and motor control.

A central observation in behavioral settings is that the calculation of cost involves a detailed knowledge of motor behavior [58, 59]. Experiments using parametric

manipulations of costs (e.g. number of level presses) and benefits (e.g. food quantity) have shown that the choices are based on a rational ordering of actions (as measured by percentages of choice and latencies; [21]). The model also accounts for this aspect as decision is based on an exact estimation of the actual effort of tested actions derived from a complete planning process.

The study of Dean et al. [57] provides indirect evidence for the proposed decision process. In this study, subjects performed rapid arm movements to hit a rewarded target. As the reward value decayed with time (a manipulation imposed by the experimenter) and movement accuracy improved with time (natural speed/accuracy relationship), the subjects had to choose a movement duration corresponding to a trade-off between reward and accuracy (see Fig. 3 in [57]). The process described in Fig. 1A is similar, but exploits the control cost (effort) rather than the movement accuracy. This is not a critical difference since there exists an univocal relationship between effort and variability [30]. Interestingly, Dean et al. [57] observed that a majority of subjects behaved optimally in this task, i.e. chose movement durations that maximized their expected gains. These results indicate that our hypothesized optimal decision process is a feasible operation for the brain.

Motor control

A central property of the model is motor control, i.e. the formation of trajectories for redundant biomechanical systems. This property is inherited from a close proximity with previous models based on optimal feedback control [10, 30]. A main novelty of this approach is to define a motor goal as a rewarded state rather than as a spatiotemporal constraint. Accordingly, movement duration is not a parameter, but an emerging characteristic of the interaction between a control policy, a controlled object,

and unexpected events (noise, perturbations). The control policy makes no difference between a *normal* and a *perturbed* state, and always elaborates commands according to the same principle. This means that a perturbation requires neither an artificial updating of movement duration [29], nor a dual control process for early (anticipatory feedforward), and late (impedance-based) motor commands [32, 60].

Interpreting the role of parameters

The model is governed by task and internal parameters that specify choices in cost/benefit situations, and kinematics and precision in motor tasks. These parameters have a psychological and neural dimension that we discuss below.

Parameter *r* reflects the well-documented influence of reward magnitude on decision making and intensity of action [61-64]. Although the observed effects are primarily mediated by physical objects (e.g. food), they can occur in the absence of reward [65], and are influenced by numerous elements. Experimental manipulations of DA transmission have been shown to bias decision making in cost/benefit situations [2, 6, 53], and alter movement intensity [2, 66]. The model offers two interpretations of these observations and of the role of DA in decision making and action, based on parameters ρ and ε (change in the perceived value of rewards or efforts). As ρ and ε have a symmetrical role, the model cannot help to decide between these interpretations. Recent studies tend to favor a relationship between effort and dopamine [19, 20, 22]. A link between ε and DA would provide a normative explanation of the strong sensitivity to response costs with preserved primary motivation for rewards following reduction of DA function [20]. Yet, the situation is probably more complex since dopamine is also involved in the valuation of reward in the absence of effort [21, 67]. Overall our results suggest that ρ and ε, through the vigor factor ρ/ε, are related to the modulation of

motivational influences. Niv et al. [51] proposed the very similar idea that tonic dopamine modulates the effort to invest in a (free operant) behavior. In contrast with our work, they focused on the rate of responding irrespective of the content of the actions, i.e. motor production. The two models are grounded on the same theoretical framework, and could complementarily help to explain the dual role of dopamine in motor behavior (e.g. vigor, time discounting) and foraging behavior (e.g. rate of reward, opportunity costs).

Parameter γ has two dimensions. On the one hand, it is a *computational* parameter that is central to the infinite-horizon formulation of optimal control [17]. On the other hand, it is a psychological parameter which is widely used in behavioral ecology and economics to represent the process by which delayed reinforcers lose value [23]. What is the status of γ in the model? Two aspects need to be elucidated. First, are three parameters (ρ, ε, γ) necessary to control movement duration? Second, is γ similar to a discount factor in behavioral economics? The first question could amount to show that γ is related to nonmotivational influences. Many elements affect movement duration, such as task instructions (e.g. move accurately; [68, 69]), task difficulty [70], and task conditions (e.g. externally-triggered movements are faster than internally-triggered movements; [71-73]). Although it might seem clear that motivational influences are not involved in these cases, it is not easy to prove it explicitly. In this framework, the latter contrast between externally- and internally-triggered movements is especially interesting. On the one hand, this contrast is similar in normal subjects and Parkinsonian patients, both on- and off-medication [71, 72]. On the other hand, Parkinsonian patients fail to properly translate motivation into action [19, 74]. The extreme case of apathetic patients is particularly revealing as they are insensitive to incentives [74] while having

"relatively spared externally-driven responses" [75]. This dichotomy is likely related to the specific implication of DA transmission in internally-generated actions [76]. Overall these results indicate that action can be modulated by influences which are independent of dopamine and motivation. The discount factor γ could mediate one of these influences.

The second question is related to the relationship between delay discounting and velocity. The study of Stevens et al. [3] is relevant to this issue. They compared the behavior of monkeys on an intertemporal choice task (a small food reward available immediately vs a delayed larger reward) and a spatial discounting choice task (a small, close reward vs a larger, more distant reward). They found that marmosets preferred larger delayed rewards in the former task, and closer, smaller rewards in the latter task. Thus their patience to wait to obtain a reward was not predictive of their will to travel farther away and for a longer time to get a larger reward. Furthermore, their travel time to the reward was not determined by their temporal discounting factor. These results indicate that decision for action is not directly governed by a discounting of time. This view is supported by neuroanatomical and neuropharmacological dissociations between effort and delay discounting in rats [2, 77]. Accordingly, the cost of time as used in the present model and in previous models [48, 49, 50, 51], seems unlike a classic temporal discounting factor, and could be specific to cost/benefit situations and motor control. This issue questions the uniqueness of time discounting across situations [50]. At odds with classical economics theories, it highlights the potential complexity and pervasiveness of the neural processes underlying computation of the cost of time [78, 79].

The model was applied to pure motor tasks in which there was no explicit reward [29, 32]. Yet, although these tasks do not apparently correspond to cost-benefit situations, there is strong experimental evidence that their execution can be modulated by cost- and benefit-related factors, e.g. loads [80], fatigue [81], task difficulty [70], attractiveness [82]. These observations suggest that pure motor tasks and reward-related motor tasks could share the same underlying representation.

Neural architecture

The model is built on a classic control/estimation architecture (Fig. 1B), which has been thoroughly discussed in the literature [83]. There is evidence that the control process is subserved by motor cortical regions [84, 85], and the estimation process by the cerebellum [86]. A central component of the model is the translation of the task parameters into a duration, a process which involves an integration of the internal parameters to calibrate costs and benefits. As discussed above, the basal ganglia and dopamine should play a crucial role in this process. In this framework, the basal ganglia would render decision making and motor control pervious to fundamental behavioral attributes (e.g. motivation, emotion, ...; [74, 87, 88]). This view is supported by studies which show that interruption of basal ganglia outputs leads to basically preserved functions [89], but deficits in behavioral modulation [74, 90].

Testing the theory

A central proposal of the model is a common basis for decision and action. The only available data that quantitatively support this proposal are those of Stevens et al. [3], which describe both choices and displacement characteristics in a spatial discounting task (Fig. 2A). In fact, any cost/benefit decision task (e.g. T-maze; [52]) could be used

to test the theory if data on displacement duration were available. As in [3], there should be a univocal relationship between displacement characteristics and choice behavior. A failure to observe this relationship would falsify the model. This would in fact correspond to a self-contradictory behavior: the costs and benefits that are estimated at the time of the decision would not be equal to those effectively encountered (during and after the movement). It should be noted that this failure would not be of the same nature as that usually reported in the field of decision making (e.g. a deviation from the laws of probabilities).

The preceding results involved locomotor patterns, but appropriate data for arm movements could be obtained using methods described in [59]. In a different domain, the model suggests that movement intensity can be modulated by nonmotivational elements, represented by the discount factor γ. One element could be urgency [71, 72]. This issue could be tested in apathetic patients, who should show a preserved sensitivity to urgency despite a loss of sensitivity to incentives [74].

Materials and Methods

Our objective is to formulate a unified model of decision making and motor control. Classical normative approaches formalize decision making as a maximization process on a *utility function* [91], and motor control as a minimization process on a *cost function* [92]. Our proposal is to build a global *utility minus cost* function (that we call again a utility function) that could govern choices and commands in a unitary way. The central issue is time, because costs in motor control are a function of time (i.e. slower movements are less costly than faster movements), as are rewards due to a discounting effect (i.e. late rewards are less valuable than immediate ones). This means that a

rational choice between two actions should involve an evaluation of their durations. However, the duration of the chosen action is only a prospective duration, valid at a given time, based on the assumption that current conditions will not change until the end of movement. The actual duration of the action can differ from this prospective duration if unexpected perturbations are encountered during the course of its execution.

We have arbitrarily chosen the notations of control theory (J for utility/cost function, u for control) rather than those of decision theory (U for utility function, a for action).

The principles of the model are first explained on a simple, deterministic example. Then the complete, stochastic version is described. The model is cast in the framework of reinforcement learning although we only exploit the optimal planning/decision processes of RL, but not the learning processes. The rationale for this choice is the following. Formally, the model corresponds to an infinite-horizon optimal control problem [93]. This jargon is typically used in economics [94], but is much less familiar in the fields of motor control and decision making, which describe similar problems in the terminology of RL [15, 16]. Furthermore, the RL framework encompasses learning processes which could explain how the proposed operations are learned by the nervous system [95, 96].

A starting example

We consider an inertial point (controlled object) described by its mass m and its state $\mathbf{x} = (p,v)$ (where p and v are the position and velocity of the object; **bold** is for vectors). This object can move along a line, actuated by a force generating system (e.g. a set of muscles). The force generating system is defined by a function h which translates a control vector $\mathbf{u}$ into muscular force ([97]; h needs not be specified for the moment).

This is a simplistic case to address e.g. the control of unidimensional saccades or single joint movements [10, 49]. The dynamics of the point is given by the general equation

$$\mathrm{d}\mathbf{x}/\mathrm{d}t = f(\mathbf{x}(t),\mathbf{u}(t)), \qquad \text{(Eq. 1)}$$

corresponding to

$$\mathrm{d}p/\mathrm{d}t = v(t)$$
$$\mathrm{d}v/\mathrm{d}t = h(u(t))/m, \qquad \text{(Eq. 2)}$$

in the case of a single muscle. To control this object means finding a *control policy*, i.e. a function $\mathbf{u}(t)$ ($t \in [t_0;t_f]$) that can displace the point between given states in the duration $t_f - t_0$. In the framework of the optimal control theory, the control policy is derived from the constraint to minimize a cost function

$$J(\mathbf{x}(t)) = \int_{[t;t_f]} L[\mathbf{x}(s),\mathbf{u}(s)]\ \mathrm{d}s, \qquad \text{(Eq. 3)}$$

for any time $t \in [t_0;t_f]$, where L is a function which generally penalizes large controls (*effort*) and deviations from a goal state (*error*; see [92] for a review). This formulation is appropriate to solve the problem of motor control, i.e. the mastering of the dynamics of articulated mechanical systems [10], but does not directly apply to a foraging problem (as described in the **Introduction**) for at least two reasons. First, function L is not concerned with values in the environment, although this difficulty could be relieved by the addition of a value-related term. Second, and more fundamental, the objective function cannot be used to specify the duration of an action, or to attribute a value to an action independent of its duration. Thus J cannot be considered as a utility function for decision making among multiple actions.

An alternative approach has been elaborated as an extension of RL in continuous time and space [16]. In this case, an infinite-horizon formulation is used where the

error/effort cost function is replaced by a time-discounted, reward/effort function (to be maximized in this case)

$$J_{\infty}(\mathbf{x}(t)) = \int_{[t;\infty]} e^{-(s-t)/\gamma} R[\mathbf{x}(s),\mathbf{u}(s)]\, ds, \qquad \text{(Eq. 4)}$$

where R is a function which weights rewarding states positively and effort negatively, and γ a time constant for discounting reward and effort. As for Eq. 3, Eq. 4 gives a recipe to find an optimal control policy $\mathbf{u}(s)$ for $s \in [t;\infty]$. For clarity, we use the symbol γ for the discount parameter as usually found in RL studies [15]. Yet the range for the discount factor is [0;1] for discrete RL, and $[0;+\infty[$ for the continuous-time formulation used here (see [16] for a correspondence between discrete and continuous RL). As in RL, a small value of γ corresponds to a large discounting effect.

We consider the case of a simple *reward minus effort* function where there is a single reward of value r at state $\mathbf{x}^*$, i.e.

$$R[\mathbf{x}(s),\mathbf{u}(s)] = \rho r \delta(||\mathbf{x}(s)-\mathbf{x}^*||) - \varepsilon||\mathbf{u}(s)||^2 \qquad \text{(Eq. 5)}$$

where δ is the function which is 1 when $\mathbf{x}(s) = \mathbf{x}^*$, and 0 everywhere else, and ρ and ε are scaling factors for reward and effort, respectively (see **Results** for a complete description of the parameters). If the inertial point starts to move at time t, reaches the rewarded state at an unknown time T, and the reward is given for a single timestep, we can write from Eq. 4 and Eq. 5, using the fact that $\mathbf{u}(s) = 0$ for $s > T$ (the point stays indefinitely at the rewarded state)

$$\begin{aligned} J_{\infty}(\mathbf{x}(t)) &= \int_{[t;\infty]} e^{-(s-t)/\gamma} [\rho r \delta(||\mathbf{x}(s)-\mathbf{x}^*||)-\varepsilon||\mathbf{u}(s)||^2]\, ds \\ &= e^{t/\gamma} \Big[\int_{[t;\infty]} e^{-s/\gamma} \rho r \delta(||\mathbf{x}(s)-\mathbf{x}^*||) ds \\ &\qquad - \varepsilon \int_{[t;T]} e^{-s/\gamma} ||\mathbf{u}(s)||^2 ds\Big] \\ &\propto \rho r e^{-T/\gamma} - \varepsilon J_{\mathbf{u}}(\mathbf{x}(t)), \qquad \text{(Eq. 6)} \end{aligned}$$

where the term $\rho r e^{-T/\gamma}$ is the discounted reward (this result comes from the fact that $\int g(s)\delta(s)ds = g(0)$ for any function g), and $J_{\mathbf{u}}(\mathbf{x}(t))$ is the motor cost

$$J_{\mathbf{u}}(\mathbf{x}(t)) = \int_{[t;T]} e^{-s/\gamma} \, ||\mathbf{u}(s)||^2 \, ds. \qquad \text{(Eq. 7)}$$

We have removed the term $\exp(t/\gamma)$ which has no influence on the maximization process. This point highlights the fact that the maximization process does not depend on current time t. For clarity, in the following, J_∞ and $J_{\mathbf{u}}$ are considered as functions of the reward time T.

The purpose of Eq. 6 is, as for Eq. 3, to obtain an optimal control policy. Maximizing J_∞ requires to find a time T and an optimal control policy $\mathbf{u}(s)$ for $s \in [t;T]$ that provide the best compromise between the *discounted reward* ($\rho r e^{-T/\gamma}$) and the *effort* ($J_{\mathbf{u}}$). This point is illustrated in Fig. 1A. Both the discounted reward and the effort ($-J_{\mathbf{u}}$ is depicted) decreases with T (i.e. a faster movement involves more effort, but leads to a less discounted reward while a slower movement takes less effort, but incurs a larger discount), and their difference takes a maximum value at a time T^* (*optimal duration*). For each T, the control policy is optimal, and is obtained by solving a classic *finite-horizon* optimal control problem with the boundary condition $\mathbf{x}(T) = \mathbf{x}^*$ ([98, 99]; see below). We note that T^* may not exist in general, depending on the shape of the reward and effort terms (Fig. 1A). Yet, this situation was never encountered in the simulations. The search of an optimal duration can be viewed both as a decision-making process (decide what is the best movement duration T^* if it exists), and a control process (if T^* exists, act with the optimal control policy defined by T^*). In the following, the maximal value of J_∞ (for $T = T^*$) will be called *utility*.

This description in terms of duration should not hide the fact that duration is only an intermediate quantity in the maximization of the utility function, and direct computation of choices and commands is possible without explicit calculus of duration [95, 96].

If there are multiple reward states in the environment, the utility defines a normative priority order among these states. A decision process which selects the action with the highest utility will choose the best possible cost/benefit compromise.

The proposed objective function involves two elements that are central to a decision making process: the benefit and the cost associated with a choice. A third element is uncertainty on the outcome of a choice. In the case where uncertainty can be represented by a probability (risk), this element could be integrated in the decision process without substantial modification of the model. A solution is to weight the reward value by the probability, in order to obtain an "expected value". Another solution is to consider that temporal discounting already contains a representation of risk [100].

In summary, equations (4) and (5) are interesting for four reasons: 1. Movement duration emerges as a compromise between discounted reward and effort; 2. The objective function is a criterion for decision-making either between different movement durations, or between different courses of action if there are multiple goals in the environment; 3. The objective function subserves both decision and control, which makes them naturally consistent. The utility that governs a decision is exactly the one that is obtained following the execution of the selected action (in the absence of noise and perturbations); 4. The objective function does not depend explicitly on time, which leads to a stationary control policy [16, 17].

General framework

For any dynamics (Eq. 1), the problem defined by Eqs. 4 and 5 is a generic infinite-horizon optimal control problem that leads, for each initial state, to an optimal movement duration and an optimal control policy (see above). This policy is also an optimal feedback control policy for each estimated state derived from an optimal state estimator [10, 99, 101, 102]. Thus the current framework is appropriate to study online movement control in the presence of noise and uncertainty. The only difference with previous approaches based on optimal feedback control [10, 99] is that movement duration is not given *a priori*, but calculated at each time to maximize an objective function.

The general control architecture is depicted in Fig. 1B. As it has been thoroughly described previously [30, 98, 99, 103], we only give here a rapid outline. The architecture contains: 1. A *controlled object* whose dynamics is described by Eq. 1, and is corrupted by noise $\mathbf{n}_{OBJ}$; 2. A *controller* defined as

$$\mathbf{u} = \mathbf{u}(\mathbf{x}^*, r, \rho, \varepsilon, \gamma, \hat{\mathbf{x}}, f), \qquad \text{(Eq. 8)}$$

which is an optimal feedback controller for Eqs. 1, 4, 5, where $\hat{\mathbf{x}}$ is the state estimate (described below); 3. An *optimal state estimator* that combines commands and sensory feedback to obtain a state estimate $\hat{\mathbf{x}}$ according to

$$d\hat{\mathbf{x}}/dt = f(\hat{\mathbf{x}}(t), \mathbf{u}(t)) + \mathbf{K}(t)[\mathbf{y}(t) - \mathbf{H}\hat{\mathbf{x}}(t-\Delta)], \qquad \text{(Eq. 9)}$$

where $\mathbf{K}$ is the Kalman gain matrix [constructed to provide an optimal weighting between the output of the forward model (first term in the rhs of Eq. 9), and the correction based on delayed sensory feedback (second term in the rhs of Eq. 9)], $\mathbf{H}$ the observation matrix, $\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t-\Delta) + \mathbf{n}_{OBS}$ the observation vector corrupted by

observation noise, and Δ the time delay in sensory feedback pathways. The observed states were the position and velocity of the controlled object.

Object noise was a multiplicative (signal-dependent) noise with standard deviation σ_{SDNm}, and observation noise was an additive (signal-independent) noise with standard deviation σ_{SINs} [98]. The rationale for this choice is to consider the simplest noisy environment: 1. Signal-dependent noise on object dynamics is necessary for optimal feedback control to implement a minimum intervention principle [10, 99]; 2. Signal-independent noise on observation is the simplest form of noise on sensory feedback. We note that a stochastic formulation was necessary to the specification of the state estimator even though most simulations actually did not involve noise.

Simulations

A simulation consisted in calculating the utility (maximal value of the objective function), and the timecourse of object state and controls for a given dynamics *f*, initial state, and parameters $\mathbf{x}^*$, r, ρ, ε, γ, σ_{SINs}, σ_{SDNm}, Δ. The solution was calculated iteratively at discretized times (timestep η). At each time t, a control policy was obtained for the current state estimate $\hat{\mathbf{x}}$ (Eq. 8). Two types of method were necessary. First, the integral term in the rhs of Eq. 6 (Eq. 7) required to solve a finite-horizon optimal control problem. This problem was solved analytically in the linear case, and numerically in the nonlinear case (see below). Second, optimal movement duration was obtained from Eq. 6 using a golden section search method [104]. Then Eqs. 1 and 9 were integrated between t and $t+\eta$ for the selected control policy and current noise levels (σ_{SINs}, σ_{SDNm}) to obtain $\mathbf{x}(t+\eta)$ and $\hat{\mathbf{x}}(t+\eta)$. The duration of the simulation was set empirically to be long enough to guarantee that the movement was completely unfolded. Actual

movement duration (and the corresponding endpoint) was determined from the velocity profile using a threshold (3 cm/s).

Three types of object were considered, corresponding to different purposes. The rationale was to use the simplest object which is deemed sufficient for the intended demonstration. Object I was a unidimensional linear object similar to that described in the starting example. The force generating system was $h(u) = u$. This object was used for decision making in a cost/benefit situation. Object II was similar to Object I, but the force generating system was a single linear second-order filter force generator (time constant τ), i.e. the dynamics was

$$\begin{aligned} \mathrm{d}p/\mathrm{d}t &= v(t) \\ \mathrm{d}v/\mathrm{d}t &= \mathrm{g}a(t)/m \\ \tau\ \mathrm{d}a/\mathrm{d}t &= -a(t) + e(t) \\ \tau\ \mathrm{d}e/\mathrm{d}t &= -e(t) + u(t), \end{aligned} \qquad \text{(Eq. 10)}$$

where a and e are muscle activation and excitation, respectively, and $\mathrm{g} = 1$ a conversion factor from activation to force. The filtering process is a minimalist analog of a muscle input/output function [105]. This object was used to study motor control in the presence of noise (relationship between amplitude, duration, and variability) [10, 30, 45]. In this case, variability was calculated as the 95% confidence interval of endpoint distribution over repeated trials ($N = 200$). Object III (IIIa and IIIb) was a classic two-joint planar arm (shoulder/elbow) actuated by two pairs of antagonist muscles. The muscles were described as nonlinear second-order filter force generators. All the details are found below. This object was used to assess characteristics of motor control in realistic motor tasks.

Parameters

For Objects I and II, the mass m was arbitrarily chosen to be 1 kg (no influence on the reported results). For Object III, the biomechanical parameters are given below. Other fixed parameters were: $\tau = 0.04$ s, $\Delta = 0.13$ s, $\eta = 0.001$ s. Noise parameters (σ_{SINs}, σ_{SDNm}) were chosen to obtain an appropriate functioning of the Kalman filter, and a realistic level of variability. The remaining parameters ($\mathbf{x}^*$, r, ρ, ε, γ) are "true" parameters that are varied to explore the model (see **Results**).

Model of the two-joint planar arm

Object III is a two-joint (shoulder, elbow) planar arm. Its dynamics is given by

$$d^2\theta/dt^2 = \mathbf{M}(\theta)^{-1}[\mathbf{T}(t) - \mathbf{C}(\theta, d\theta/dt)d\theta/dt],$$

where $\theta = (\theta_1, \theta_2)$ is the vector of joint angles, $\mathbf{M}$ the inertia matrix, $\mathbf{C}$ the matrix of velocity-dependent forces, $\mathbf{W}$ an optional velocity-dependent force field matrix, and $\mathbf{T}(t)$ the vector of muscle torques defined by

$$\mathbf{T}(t) = \mathbf{A}\mathbf{F}_{max}[\mathbf{a}(t)]^+,$$

where $\mathbf{A}$ is the matrix of moment arms, $\mathbf{F}_{max}$ the matrix of maximal muscular forces, and $\mathbf{a}$ the vector of muscular activations resulting from the application of a control signal $\mathbf{u}(t)$ (see Eq. 10).

For each segment (1: upper arm, 2: forearm), l is the length, I the inertia, m the mass, and c the distance to center of mass from the preceding joint. Matrix $\mathbf{M}$ is

$$[M_{11}\ M_{12}\ ;\ M_{21}\ M_{22}\ ],$$

with

$$M_{11} = I_1 + I_2 + m_1 c_1^2 + m_2(l_1^2 + c_2^2 + 2 l_1 c_2 \cos(\theta_2))$$

$$M_{12} = M_{21} = I_2 + m_2(c_2^2 + l_1 c_2 \cos(\theta_2))$$

$$M_{22} = I_2 + m_2 c_2^2$$

Matrix **C** is

$$[C_{11}\ C_{12}\ ;\ C_{21}\ C_{22}\ ],$$

with

$$C_{11} = -\ m_2\ l_1 c_2 \sin(\theta_2) d\theta_2/dt - 0.05$$

$$C_{12} = -\ m_2\ l_1 c_2 \sin(\theta_2)(d\theta_1/dt + d\theta_2/dt) - 0.025$$

$$C_{21} = m_2\ l_1 c_2 \sin(\theta_2) d\theta_1/dt - 0.025$$

$$C_{22} = -\ 0.05$$

Matrix **W** is $\mathbf{JDJ}^T$, where **J** is the Jacobian matrix of the arm, and **D** (Ns/m) is

$$[\ -10.1\ -11.2\ ;\ -11.2\ 11.1\ ].$$

Matrix $\mathbf{F}_{max}$ (N) is diag([700;382;572;449]). Matrix **A** (m) is

$$[\ .04\ -.04\ 0\ 0\ ;\ 0\ 0\ .025\ -.025\ ].$$

Two sets of parameter values were used in the simulations. For Object IIIa, we used the values found in [29] (in S.I.): $l_1 = .30$, $l_2 = .33$, $I_1 = .025$, $I_2 = .045$, $m_1 = 1.4$, $m_2 = 1.1$, $c_1 = .11$, $c_2 = .16$. For Object IIIb, we used the values given in [32]: $l_1 = .33$, $l_2 = .34$, $I_1 = .0141$, $I_2 = .0188$, $m_1 = 1.93$, $m_2 = 1.52$, $c_1 = .165$, $c_2 = .19$.

Resolution of the optimal control problem

The problem is to find the sequence of control $\mathbf{u}(t)$ which optimizes the objective function $J_u(T)$ (Eq. 7), and conforms to the boundary conditions $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{x}(T) = \mathbf{x}^*$ for a given dynamic f. The general approach to solve this problem is based on variational calculus [106]. The first step is to construct the Hamiltonian function which combines the objective function and the dynamic thanks to the Lagrangian multipliers (or co-state) denoted by λ

$$H(\mathbf{x},\mathbf{u},\lambda,t) = \varepsilon\mathbf{u}(t)^T\mathbf{u}(t) + \lambda(t)^T f(\mathbf{x}(t),\mathbf{u}(t)).$$

The optimal control minimizes the Hamiltonian, a property known as the Pontryagin's minimum principle given formally by

$$d\mathbf{x}/dt = \partial H/\partial\lambda = f(\mathbf{x}(t),\mathbf{u}(t)) \qquad \text{(Eq. 11)}$$

$$d\lambda/dt = -\partial H/\partial\mathbf{x} + \lambda(t)/\gamma = -\lambda(t)\partial f/\partial\mathbf{x} + \lambda(t)/\gamma \qquad \text{(Eq. 12)}$$

$$0 = \partial H/\partial\mathbf{u} = \varepsilon\mathbf{u}(t) + \lambda(t)\partial f/\partial\mathbf{u} \qquad \text{(Eq. 13)}$$

Equation (12), widely used in economics, is slightly different from what is usually used in the motor control literature because of the discounting factor in the objective function. We will thereafter consider two methods to solve this set of differential equations depending on the complexity of the dynamics.

Linear case

If the dynamic f is linear, as for Objects I and II, the system of differential equations (Eqs. 11, 12, 13) is also linear, and can be solved analytically. We rewrite the dynamics as

$$f(\mathbf{x}(t),\mathbf{u}(t)) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t).$$

From Eq. 13, we can reformulate the optimal control $\mathbf{u}^*(t)$ as

$$\mathbf{u}^*(t) = -\mathbf{B}^T\lambda(t)/\varepsilon.$$

In order to find $\lambda(t)$, we then replace $\mathbf{u}(t)$ by $\mathbf{u}^*(t)$ in Eqs. 11 and 12, and get

$$d\mathbf{x}/dt = \mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{B}^T\lambda/\varepsilon$$

$$d\lambda/dt = (-\mathbf{A}^T+\mathbf{I}/\gamma)\lambda, \qquad \text{(Eq. 14)}$$

where $\mathbf{I}$ is the identity matrix. The resolution of this system gives the optimal trajectory of the state and the co-state

$$(\mathbf{x}^* \ \lambda^*)^T = \Gamma(t)\mathbf{C},$$

where Γ is the analytic solution to Eq. 14, and $\mathbf{C}$ can be deduced from the boundary conditions [99]. Finally, we replace λ by λ^* in Eq. 14 to get the value of the optimal control. From Eq. 6, we obtain an analytic version of the utility, from which we can derive the optimal duration T^* analytically. Symbolic calculus was performed with Maxima (Maxima, a Computer Algebra System. Version 5.18.1 (2009) http://maxima.sourceforge.net/).

Nonlinear case

When the dynamics is nonlinear (Object III), the set of differential equations (Eqs. 11, 12, 13) cannot be solved directly. However, the minimum of the Hamiltonian (and thus the optimal control) can be found through numerical methods using a gradient descent method. The detail of the existing algorithms is outside the scope of this article, and the reader is referred to [101], and [106].

Acknowledgements

We thank O. Sigaud, A. Terekhov, P. Baraduc, and M. Desmurget for fruitful discussions.

References

1. Stephens DW, Krebs JR (1986) Foraging Theory. Princeton, NJ: Princeton University Press. 262 p.
2. Denk F, Walton ME, Jennings KA, Sharp T, Rushworth MF, Bannerman DM (2005) Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. Psychopharmacology (Berl) 179: 587-596.
3. Stevens JR, Rosati AG, Ross KR, Hauser MD (2005) Will travel for food: Spatial discounting in two new world monkeys. Curr Biol 15: 1855-1860.
4. Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MF (2006) Separate neural pathways process different decision costs. Nat Neurosci 9: 1161-1168.
5. Walton ME, Kennerley SW, Bannerman DM, Phillips PEM, Rushworth MF (2006) Weighing up the benefits of work: Behavioral and neural analyses of effort-related decision making. Neural Netw 19: 1302-1314.
6. Floresco SB, Tse MT, Ghods-Sharifi S (2008) Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. Neuropsychopharmacology 33: 1966-1979.
7. Braun DA, Nagengast AJ, Wolpert DM (2011) Risk-sensitivity in sensorimotor control. Front Hum Neurosci 5: 1.
8. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. Econometrica 47: 263-291.

9. Prévost C, Pessiglione M, Météreau E, Cléry-Melin ML, Dreher J-C (2010) Separate valuation subsystems for delay and effort decision costs. J Neurosci 30: 14080-14090.

10. Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nat Neurosci 5: 1226-1235.

11. Guigon E, Baraduc P, Desmurget M (2007) Computational motor control: Redundancy and invariance. J Neurophysiol 97: 331-347.

12. Trommershäuser J, Maloney LT, Landy MS (2003) Statistical decision theory and rapid, goal-directed movements. J Opt Soc Am A 20: 1419-1433.

13. Trommershäuser J, Maloney LT, Landy MS (2003) Statistical decision theory and trade-offs in motor response. Spat Vis 16: 255-275.

14. Trommershäuser J, Gepshtein S, Maloney LT, Landy MS, Banks MS (2005) Optimal compensation for changes in task-relevant movement variability. J Neurosci 25: 7169-7178.

15. Sutton RS, Barto AG (1998) Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press. 322 p.

16. Doya K (2000) Reinforcement learning in continuous time and space. Neural Comput 12: 219-245.

17. Todorov E (2007) Optimal control theory. In: Doya K, Ishii S, Pouget A, Rao RPN, editors. Bayesian Brain: Probabilistic Approaches to Neural Coding. Cambridge, MA: MIT Press. pp. 269-298.

18. Körding K (2007) Decision theory: What "should" the nervous system do? Science 318: 606-610.

19. Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27: 7105-7116.

20. Salamone JD, Correa M, Farrar A, Mingote SM (2007) Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. Psychopharmacology (Berl) 191: 461-482.

21. Gan JO, Walton ME, Phillips PE (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13: 25-27.

22. Kurniawan IT, Guitart-Masip M, Dolan RJ (2011) Dopamine and effort-based decision making. Front Neurosci 5: 81.

23. Green L, Myerson J (1996) Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting times. Am Zool 36: 496-505.

24. Decety J, Jeannerod M, Prablanc C (1989) The timing of mentally represented actions. Behav Brain Res 34: 35-42.

25. Bakker M, de Lange FP, Stevens JA, Toni I, Bloem BR (2007) Motor imagery of gait: A quantitative approach. Exp Brain Res 179: 497-504.

26. Hicheur H, Pham QC, Arechavaleta G, Laumond J-P, Berthoz A (2007) The formation of trajectories during goal-oriented locomotion in humans. I. A stereotyped behaviour. Eur J Neurosci 26: 2376-2390.

27. Kunz BR, Creem-Regehr SH, Thompson WB (2009) Evidence for motor simulation in imagined locomotion. J Exp Psychol: Hum Percept Perform 35: 1458-1471.

28. Mühlhoff N, Stevens JR, Reader SM (2011) Spatial discounting of food and social rewards in guppies (*Poecilia reticulata*). Front Psychology 2: 68.

29. Liu D, Todorov E (2007) Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. J Neurosci 27: 9354-9368.

30. Guigon E, Baraduc P, Desmurget M (2008) Computational motor control: Feedback and accuracy. Eur J Neurosci 27: 1003-1016.

31. Gordon J, Ghilardi MF, Cooper SE, Ghez C (1994) Accuracy of planar reaching movements. II. Systematic extent errors resulting from inertial anisotropy. Exp Brain Res 99: 112-130.

32. Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning a motor task. J Neurosci 14: 3208-3224.

33. Hikosaka O, Wurtz RH (1985) Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus. J Neurophysiol 53: 266-291.

34. Hikosaka O, Wurtz RH (1985) Modification of saccadic eye movements by GABA-related substances. II. Effect of muscimol in monkey substantia nigra pars reticulata. J Neurophysiol 53: 292-308.

35. Kato M, Miyashita N, Hikosaka O, Matsumura M, Usui S, Kori A (1995) Eye movements in monkeys with local dopamine depletion in the caudate nucleus. 1. Deficits in spontaneous saccades. J Neurosci 15: 912-927.

36. Alamy M, Pons J, Gambarelli D, Trouche E (1996) A defective control of small amplitude movements in monkeys with globus pallidus lesions: An experimental study on one component of pallidal bradykinesia. Behav Brain Res 72: 57-62.

37. Georgiou N, Phillips JG, Bradshaw JL, Cunnington R, Chiu E (1997) Impairments of movement kinematics in patients with Huntington's disease: A comparison with and without a concurrent task. Mov Disorders 12: 386-396.

38. Robichaud JA, Pfann KD, Comella CL, Corcos DM (2002) Effect of medication on EMG patterns in individuals with Parkinson's disease. Mov Disorders 17: 950-960.

39. Negrotti A, Secchi C, Gentilucci M (2005) Effects of disease progression and L-dopa therapy on the control of reaching-grasping in Parkinson's disease. Neuropsychologia 43: 450-459.

40. Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47: 381-391.

41. Bainbridge L, Sanders M (1972) The generality of Fitts's law. J Exp Psychol 96: 130-133.

42. Osu R, Kamimura N, Iwasaki H, Nakano E, Harris CM, Wada Y, Kawato M (2004) Optimal impedance control for task achievement in the presence of signal-dependent noise. J Neurophysiol 92: 1199-1215.

43. Selen LP, Beek PJ, van Dieen JH (2006) Impedance is modulated to meet accuracy demands during goal-directed arm movements. Exp Brain Res 172: 129-138.

44. Meyer DE, Abrams RA, Kornblum S, Wright CE, Smith JEK (1988) Optimality in human motor performance: Ideal control of rapid aimed movement. Psychol Rev 95: 340-370.

45. Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. Nature 394: 780-784.

46. Tanaka H, Krakauer JW, Qian N (2006) An optimization principle for determining movement duration. J Neurophysiol 95: 3875-3886.

47. Wu SW, Delgado MR, Maloney LT (2009) Economic decision-making compared with an equivalent motor task. Proc Natl Acad Sci USA 106: 6088-6093.

48. Hoff B (1994) A model of duration in normal and perturbed reaching movement. Biol Cybern 71: 481-488.

49. Harris CM, Wolpert DM (2006) The main sequence of saccades optimizes speed-accuracy trade-off. Biol Cybern 95: 21-29.

50. Shadmehr R, Orban de Xivry JJ, Xu-Wilson M, Shih TY (2010) Temporal discounting of reward and the cost of time in motor control. J Neurosci 30: 10507-10516.

51. Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: Opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191: 507-520.

52. Salamone JD, Cousins MS, Bucher S (1994) Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. Behav Brain Res 65: 221-229.

53. Kurniawan IT, Seymour B, Talmi D, Yoshida W, Chater N, Dolan RJ (2010) Choosing to make an effort: The role of striatum in signaling physical effort of a chosen action. J Neurophysiol 104: 313-321.

54. Battaglia PW, Schrater PR (2007) Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. J Neurosci 27: 6984-6994.

55. Trommershäuser J, Maloney LT, Landy MS (2008) Decision making, movement planning and statistical decision theory. Trends Cogn Sci 12: 291-297.

56. Nagengast AJ, Braun DA, Wolpert DM (2010) Risk-sensitive optimal feedback control accounts for sensorimotor behavior under uncertainty. PLoS Comput Biol 6: e1000857.

57. Dean M, Wu SW, Maloney LT (2007) Trading off speed and accuracy in rapid, goal-directed movements. J Vis 7: 10.

58. Phillips PEM, Walton ME, Jhou TC (2007) Calculating utility: Preclinical evidence for cost-benefit analysis by mesolimbic dopamine. Psychopharmacology (Berl) 191: 483-495.

59. Cos I, Bélanger N, Cisek P (2011) The influence of predicted arm biomechanics on decision making. J Neurophysiol 105: 3022-3033.

60. Bhushan N, Shadmehr R (1999) Computational nature of human adaptive control during learning of reaching movements in force fields. Biol Cybern 81: 39-60.

61. Crespi LP (1942) Quantitative variation in incentive and performance in the white rat. Am J Psychol 55: 467-517.

62. Brown VJ, Bowman EM (1995) Discriminative cues indicating reward magnitude continue to determine reaction time of rats following lesions of the nucleus accumbens. Eur J Neurosci 7: 2479-2485.

63. Watanabe K, Lauwereyns J, Hikosaka O (2003) Effects of motivational conflicts on visually elicited saccades in monkeys. Exp Brain Res 152: 361-367.

64. Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. J Neurosci 29: 13365-13376.

65. Aarts H, Custers R, Marien H (2008) Preparing and motivating behavior outside of awareness. Science 319: 1639.

66. Choi WY, Morvan C, Balsam PD, Horvitz JC (2009) Dopamine D1 and D2 antagonist effects on response likelihood and duration. Behav Neurosci 123: 1279-1287.

67. Nicola SM (2010) The flexible approach hypothesis: Unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. J Neurosci 30: 16585-16600.

68. Brown SH, Hefter H, Mertens M, Freund HJ (1990) Disturbances in human arm trajectory due to mild cerebellar dysfunction. J Neurol Neurosurg Psychiatry 53: 306-313.

69. Hefter H, Brown SH, Cooke JD, Freund HJ (1996) Basal ganglia and cerebellar impairment differentially affect the amplitude and time scaling during the performance of forearm step tracking movements. Electromyogr Clin Neurophysiol 36: 121-128.

70. Montagnini A, Chelazzi L (2005) The urgency to look: Prompt saccades to the benefit of perception. Vis Res 45: 3391-3401.

71. Majsak MJ, Kaminski TR, Gentile AM, Flanagan JR (1998) The reaching movements of patients with Parkinson's disease under self-determined maximal speed and visually cued conditions. Brain 121: 755-766.

72. Ballanger B, Thobois S, Baraduc P, Turner RS, Broussolle E, Desmurget M (2006) "Paradoxical kinesis" is not a hallmark of Parkinson's disease but a general property of the motor system. Mov Disorders 21: 1490-1495.

73. Welchman AE, Stanley J, Schomers MR, Miall RC, Bülthoff HH (2010) The quick and the dead: When reaction beats intention. Proc Biol Sci 277: 1667-1674.

74. Schmidt L, d'Arc BF, Lafargue G, Galanaud D, Czernecki V, Grabli D, Schüpbach M, Hartmann A, Lévy R, Dubois B, Pessiglione M (2008) Disconnecting force from money: Effects of basal ganglia damage on incentive motivation. Brain 131: 1303-1310.

75. Lévy R, Czernecki V (2006) Apathy and the basal ganglia. J Neurol 253: VII54-61.

76. Jahanshahi M, Frith CD (1998) Willed action and its impairment. Cogn Neuropsychol 15: 483-533.

77. Ghods-Sharifi S, Floresco SB (2010) Differential effects on effort discounting induced by inactivations of the nucleus accumbens core or shell. Behav Neurosci 124: 179-191.

78. Schweighofer N, Shishida K, Han CE, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2006) Humans can adopt optimal discounting strategy under real-time constraints. PLoS Comput Biol 2: e152.

79. Peters J, Büchel C (2011) The neural mechanisms of inter-temporal decision-making: Understanding variability. Trends Cogn Sci 15: 227-239.

80. Bock O (1990) Load compensation in human goal-directed arm movements. Behav Brain Res 41: 167-177.

81. Corcos DM, Jiang HY, Wilding J, Gottlieb GL (2002) Fatigue induced changes in phasic muscle activation patterns for fast elbow flexion movements. Exp Brain Res 142: 1-12.

82. Xu-Wilson M, Zee DS, Shadmehr R (2009) The intrinsic value of visual information affects saccade velocities. Exp Brain Res 196: 475-481.

83. Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. Exp Brain Res 185: 359-381.

84. Scott SH (2004) Optimal feedback control and the neural basis of volitional motor control. Nat Rev Neurosci 5: 532-546.

85. Guigon E, Baraduc P, Desmurget M (2007) Coding of movement- and force-related information in primate primary motor cortex: A computational approach. Eur J Neurosci 26: 250-260.

86. Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Netw 9: 1265-1279.

87. Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau H, Dolan RJ, Frith CD (2007) How the brain translates money into force: A neuroimaging study of subliminal motivation. Science 316: 904-906.

88. Schmidt L, Cléry-Melin ML, Lafargue G, Valabrègue R, Fossati P, Dubois B, Pessiglione M (2009) Get aroused and be stronger: Emotional facilitation of physical effort in the human brain. J Neurosci 29: 9450-9457.

89. Turner RS, Desmurget M (2010) Basal ganglia contributions to motor control: A vigorous tutor. Curr Opin Neurobiol 20: 704-716.

90. Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 96: 1441-1455.

91. Pratt JW, Raiffa H, Schlaifer R (1995) Introduction to Statistical Decision Theory. Cambridge: MIT Press. 895 p.

92. Todorov E (2004) Optimality principles in sensorimotor control. Nat Neurosci 7: 907-915.

93. Bertsekas DP, Shreve SE (1996) Stochastic Optimal Control: The Discrete Time Case. Belmont: Athena Scientific. 323 p.

94. Kunkel P, von Dem Hagen O (2000) Numerical solution of infinite-horizon optimal-control problems. Comput Econ 16: 189-205.

95. Simpkins A, Todorov E (2009) Practical numerical methods for stochastic optimal control of biological systems in continuous time and space. In: Proceedings of the IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning; 30 March-2 April 2009; Nashville, Tennessee, United States. ADPRL 2006. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4927547. Accessed 24 August 2012.

96. Marin D, Decock J, Rigoux L, Sigaud O (2011) Learning cost-efficient control policies with XCSF: Generalization capabilities and further improvement. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation; 12-16 July 2011; Dublin, Ireland. GECCO 2011.

97. Zajac FE (1989) Muscle and tendon: Models, scaling, and application to biomechanics and motor control. Crit Rev Biomed Eng 17: 359-415.

98. Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Comput 17: 1084-1108.

99. Guigon E, Baraduc P, Desmurget M (2008) Optimality, stochasticity, and variability in motor behavior. J Comput Neurosci 24: 57-68.

100. Platt ML, Huettel SA (2008) Risky business: The neuroeconomics of decision making under uncertainty. Nat Neurosci 11: 398-403.

101. Bryson AE, Ho Y-C (1975) Applied Optimal Control - Optimization, Estimation, and Control. New York: Hemisphere Publ Corp. 481 p.

102. Stengel RF (1986) Stochastic Optimal Control: Theory and Application. New York, NY: Wiley. 638 p.

103. Guigon E (2010) Active control of bias for the control of posture and movement. J Neurophysiol 104: 1090-1102.

104. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2002) Numerical Recipes in C. The Art of Scientific Computing (2nd ed). New York: Cambridge University Press. 994 p.

105. van der Helm FCT, Rozendaal LA (2000) Musculoskeletal systems with intrinsic and proprioceptive feedback. In: Winters JM, Crago PE, editors. Biomechanics and Neural Control of Posture and Movement. New York, NY: Springer. pp. 164-174.

106. Kirk DE (2004) Optimal Control Theory: An Introduction. Mineola, NY: Dover. 452 p.

Figure legends

Figure 1. Objective function and model architecture **A**. Objective function (*thick*) as a function of movement duration, built from the sum of a discounted reward term (*thin*) and a discounted effort term (*dashed*). Optimal duration is indicated by a vertical *dotted* line. **B**. Architecture of the infinite-horizontal optimal feedback controller. See **Text** for notations.

Figure 2. Simulation of Stevens [3]. **A**. Cost/benefit choice task between a reference option (small reward/short distance) and a test option (large reward/long distance). **B**. Utility vs distance. The dotted line indicates the utility for the reference option ($r = 1$, distance = .35 m). The solid line gives the utility for the test option ($r = 3$) for different distances (range .35-2.45 m). An arrow indicates the distance at which the preference changes. Results obtained with Object I. Parameters: $\rho/\varepsilon = 1$, $\gamma = 2$. **C**. Vigor and discount factors for synthetic monkeys (*black*: marmosets; *gray*: tamarins) derived from [3]. The figure was built in the following way. Mean m and standard deviation σ of displacement duration were obtained from Fig. 3 in [3] for each species and each amplitude. For each species, a random sample was drawn from the corresponding Gaussian distribution $N(m,\sigma)$ for each amplitude, giving two durations. These two durations were used to identify a unique pair of parameters (vigor, discount). Each point corresponds to one pair. See Text for further explanation. **D**. Indifference points corresponding to the simulated monkeys shown in **C** (T = tamarin, M = marmoset). Bold bar is the median, hinges correspond to the first and third quartile (50% of the population), and whiskers to the first and ninth decile (90% of the population).

E. Probability of choosing the large reward option according the test distance. Solid lines are the experimental data from Stevens [3]. *Dashed* lines and *shaded* areas correspond respectively to the mean and the 95% confidence interval of the decision process derived from the simulated utilities and a soft-max rule. The temperature parameter was selected for each monkey to fit empirical data.

Figure 3. Basic characteristics of motor control. **A**. Trajectories for movements of different amplitudes (direction: 45 deg; 5, 10, 15, 20, 25, 30 cm). **B**. Trajectories for movements in different directions (10 cm). **C**. Amplitude/duration scaling law and velocity profiles (inset) for the movements in **A**. **D**. Direction/duration (*plain line*), direction/apparent inertia (*dotted line*; arbitrary unit; [31]). Results obtained with Object IIIa. Initial arm position (deg): (75,75). Parameters: $r = 40$, $\rho/\varepsilon = 1/300$, $\gamma = .5$, $\sigma_{SINs} = .001$, $\sigma_{SDNm} = 1$.

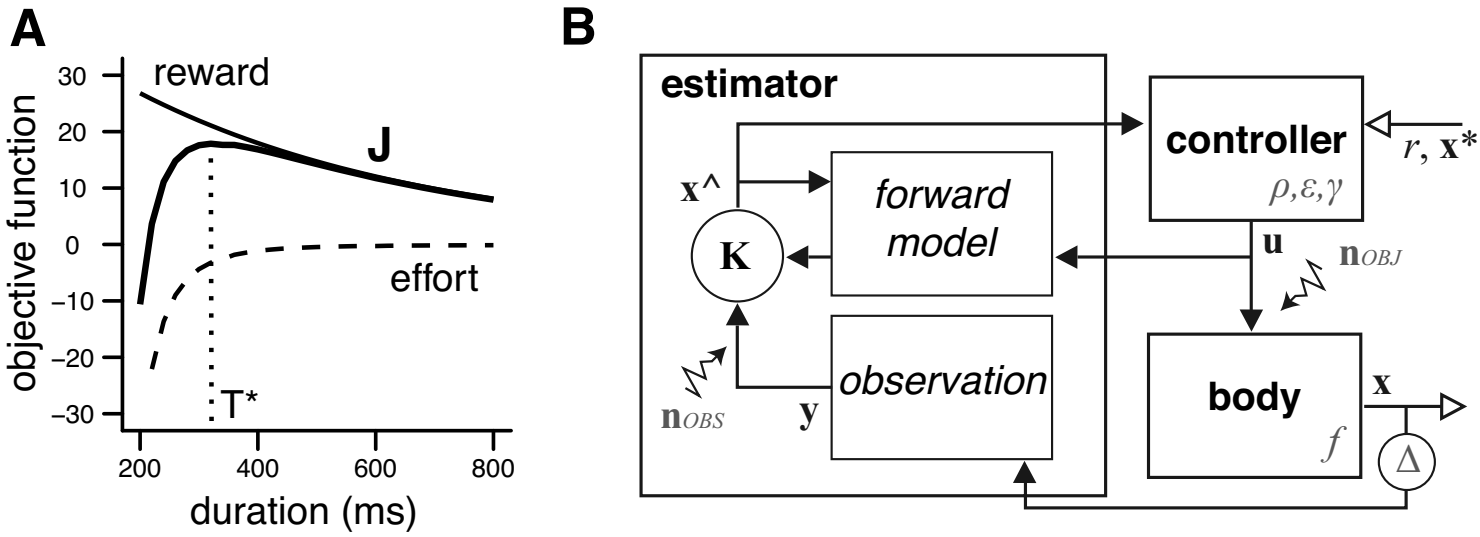
Figure 4. Simulation of Liu and Todorov [29]. **A**. Simulated trajectories for reaching movements toward a target which jumps unexpectedly up or down, 100 ms, 200 ms or 300 ms after movement onset. **B**. Corresponding velocity profiles. **C**. Arrival time as a function of the timing of the perturbation. Results obtained with Object IIIa. Initial arm position (deg): (15,120). Same parameters as in Fig. 3.

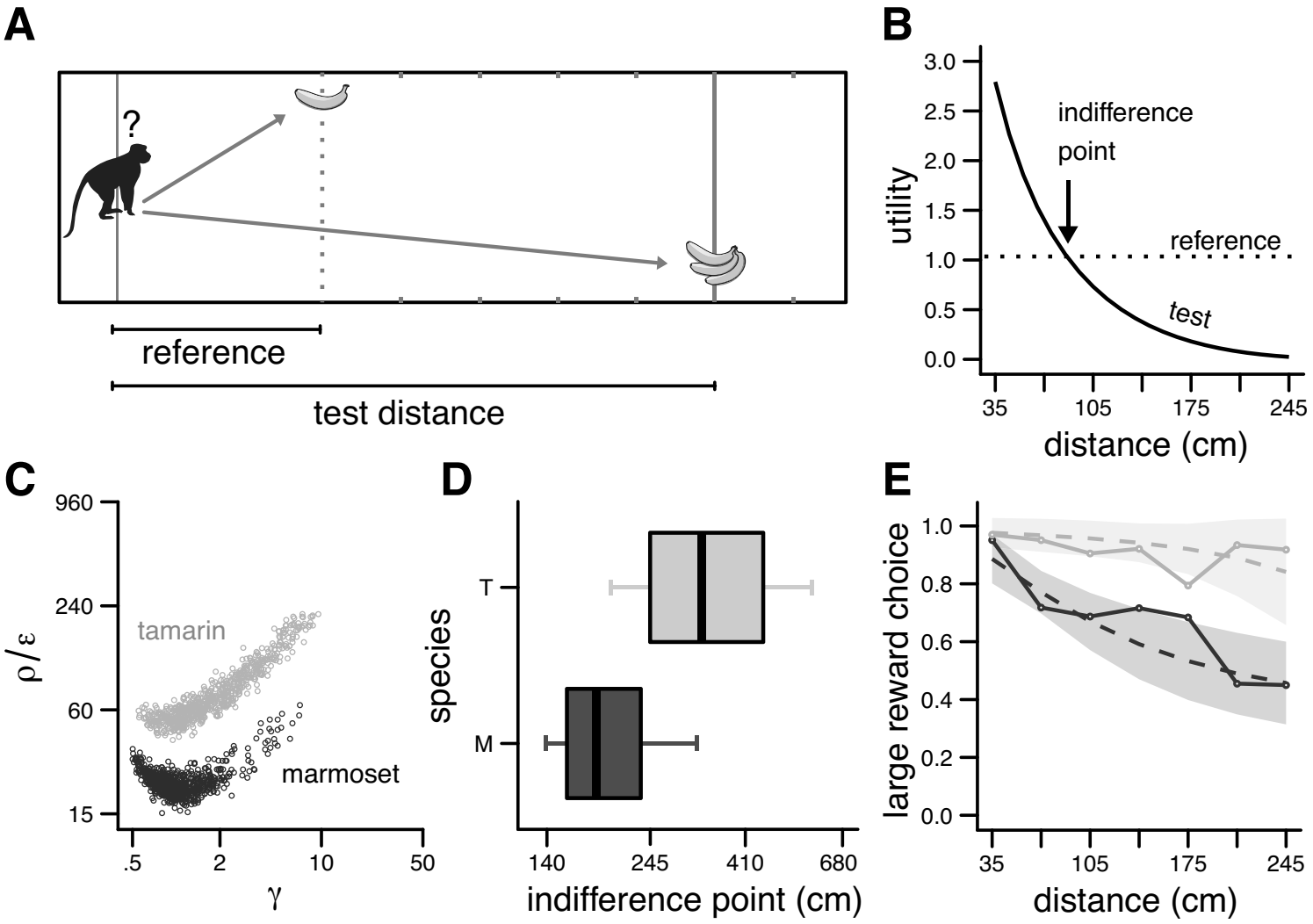
Figure 5. Simulation of Shadmehr and Mussa-Ivaldi [32]. **A**. Velocity profiles for unperturbed movements in four directions. **B**. Hand trajectories during exposure to a velocity-dependent force field. **C**. Velocity profiles for perturbed movements in four

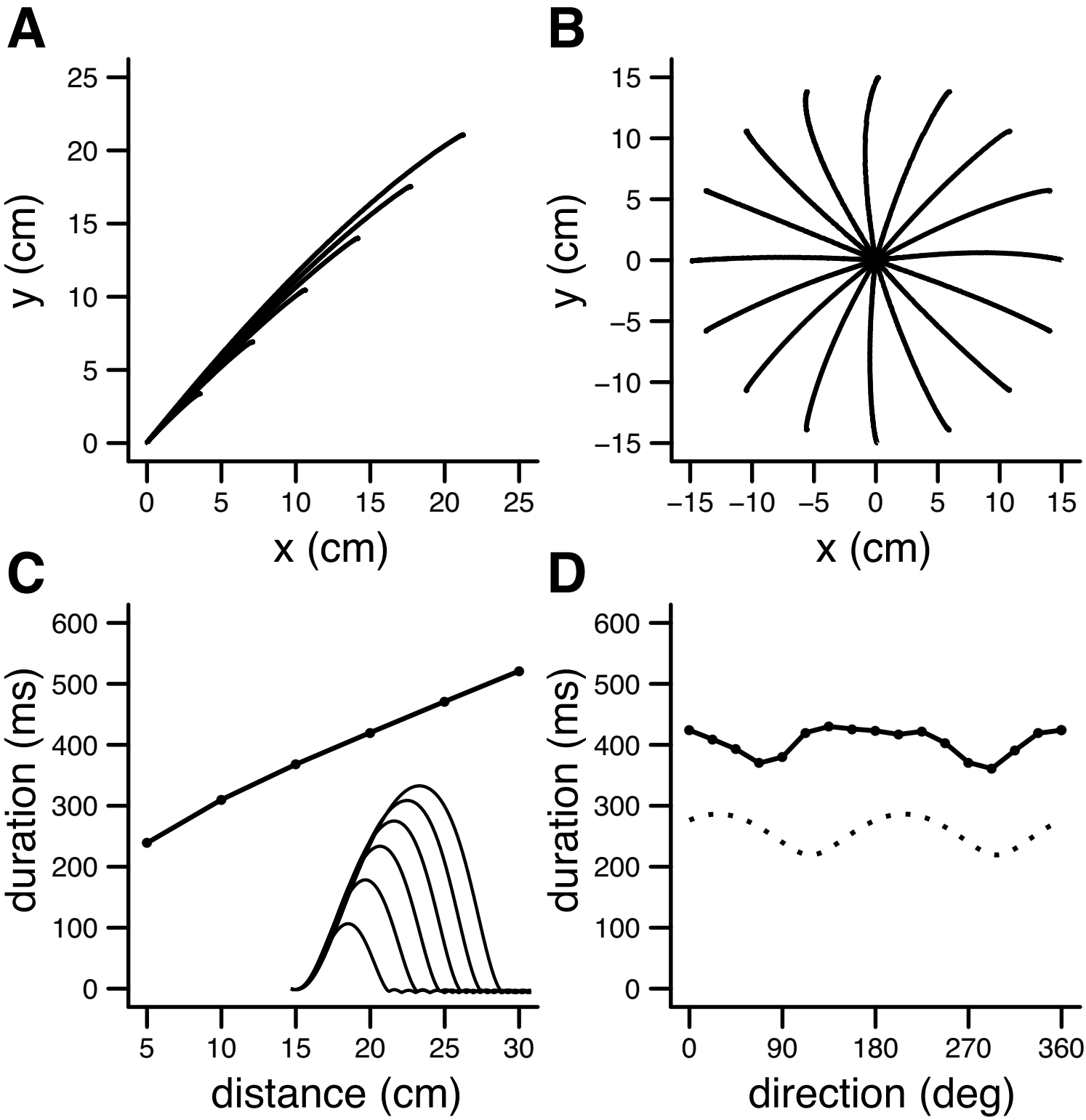
directions (data from **B**). Results obtained with Object IIIb. Initial arm position (deg): (15,100). Same parameters as in Fig. 3.

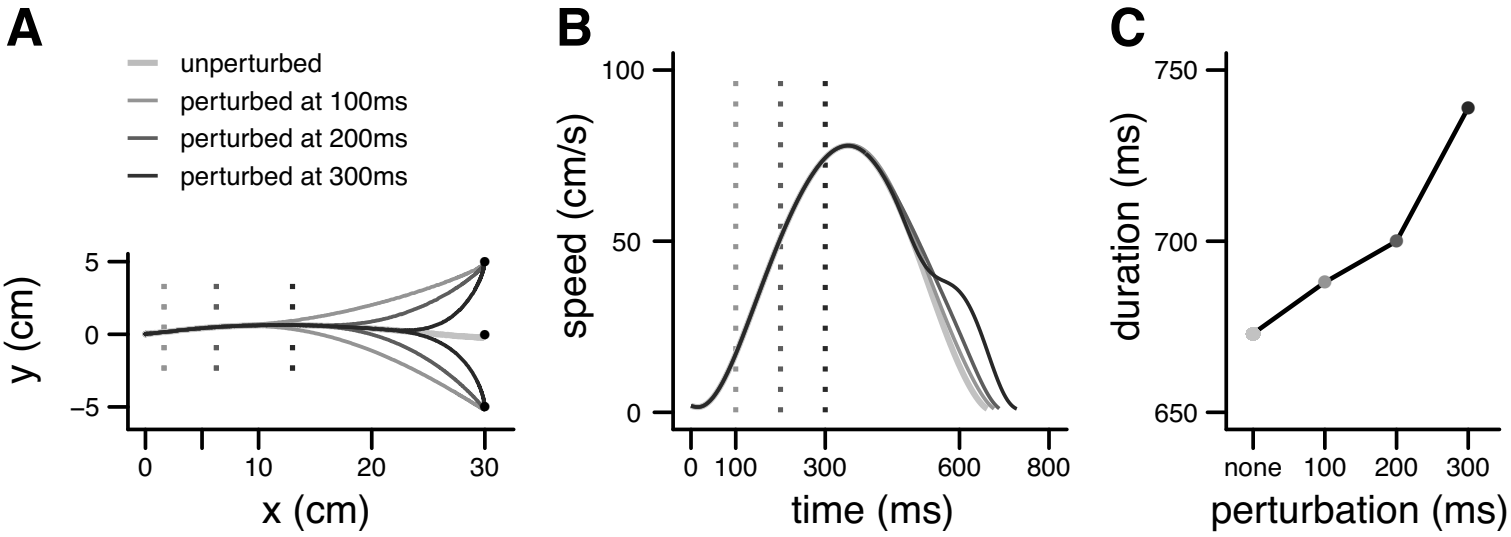
Figure 6. Influence of parameters. **A**. Change in the distance/utility relationship induced by a decrease in vigor: ρ/ε from 50 (*gray*) to 16 (*black*). Same experiment as in Fig. 2A. Parameters: $r = 1$, $\gamma = 2$. **B**. Same as **A** for a decrease in the value of discount factor: γ from 4 (*gray*) to 1 (*black*). Parameters: $r = 1$, $\rho/\varepsilon = 50$. **C**. Change in movement duration corresponding to the results in **A**. **D**. Change in movement duration corresponding to the results in **B**. Results obtained with Object I.

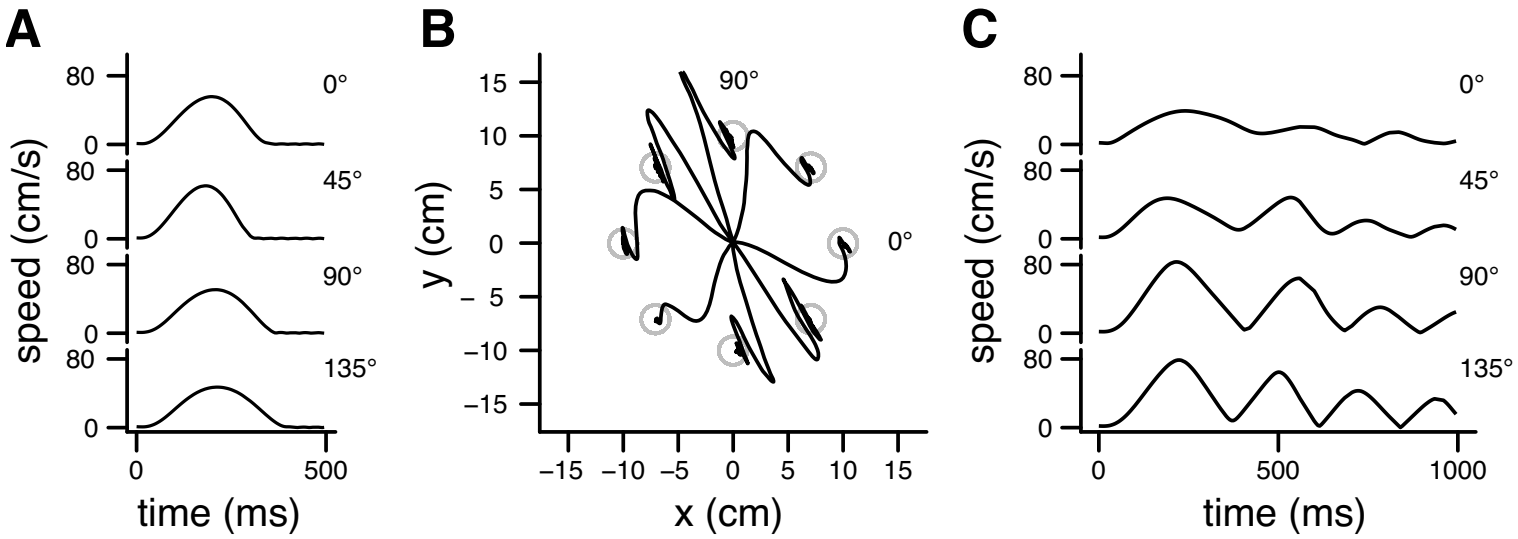
Figure 7. Fitts' law and variability. **A**. Duration as a function of the index of difficulty (ID) for 3 distances (10, 20 and 30 cm) and different values of vigor and discount (see legend). **B**. Typical spatiotemporal variability (s.d. of position). **C**. Endpoint variability for different values of the discount factor. Color is for the level of vigor (legend in **A**). Results obtained with Object II. Parameters: distance = 30 cm, $r = 1$, $\rho/\varepsilon = 100$, $\gamma = 2$, $\sigma_{SINs} = .001$, $\sigma_{SDNm} = 1$.

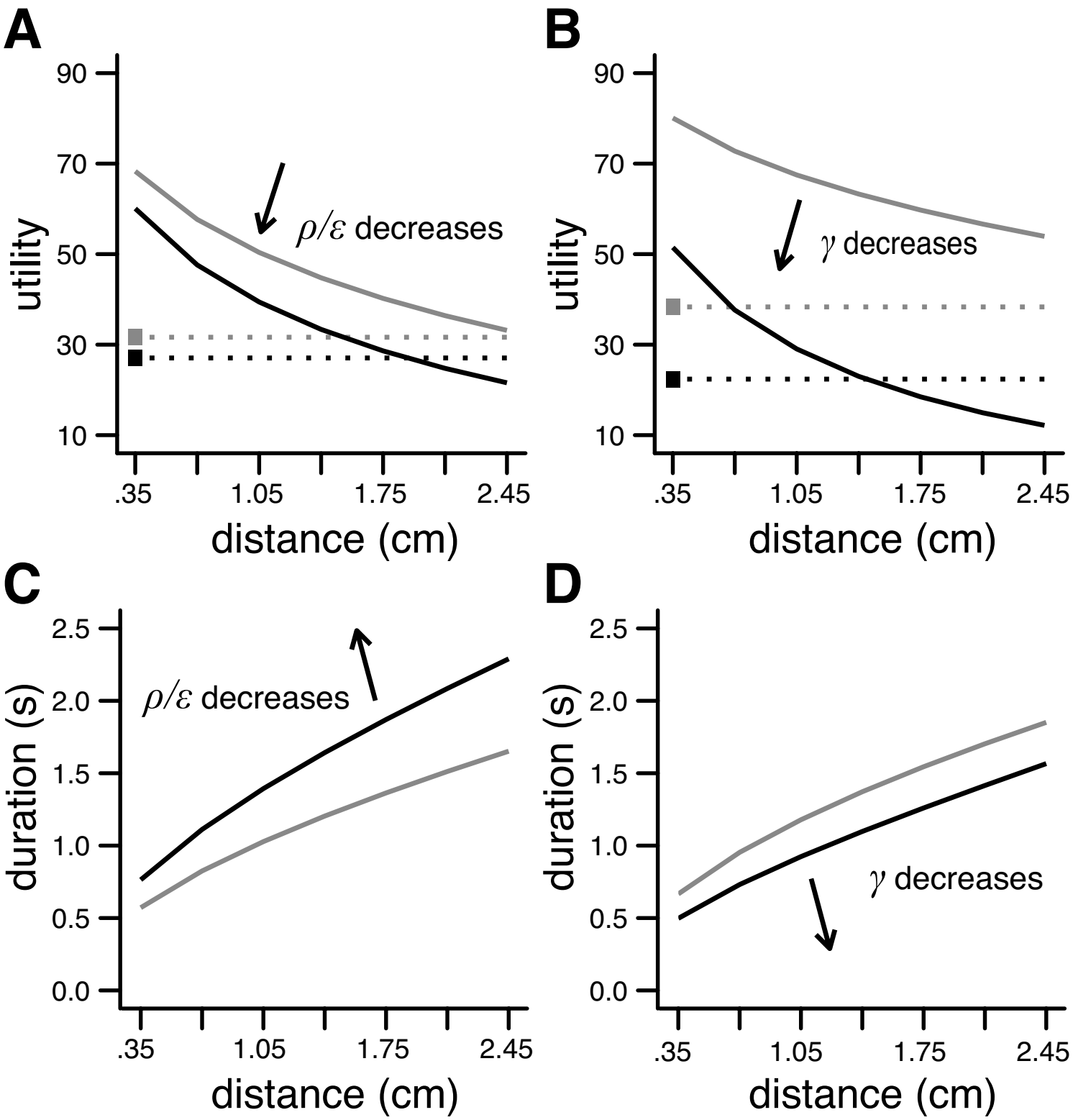

A
reward
J
effort
T*
objective function
duration (ms)
30
20
10
0
−10
−20
−30
200
400
600
800
B
estimator
controller
ρ,ε,γ
r, x*
x^
K
forward model
u
n_{OBJ}
n_{OBS}
y
observation
body
f
x
Δ

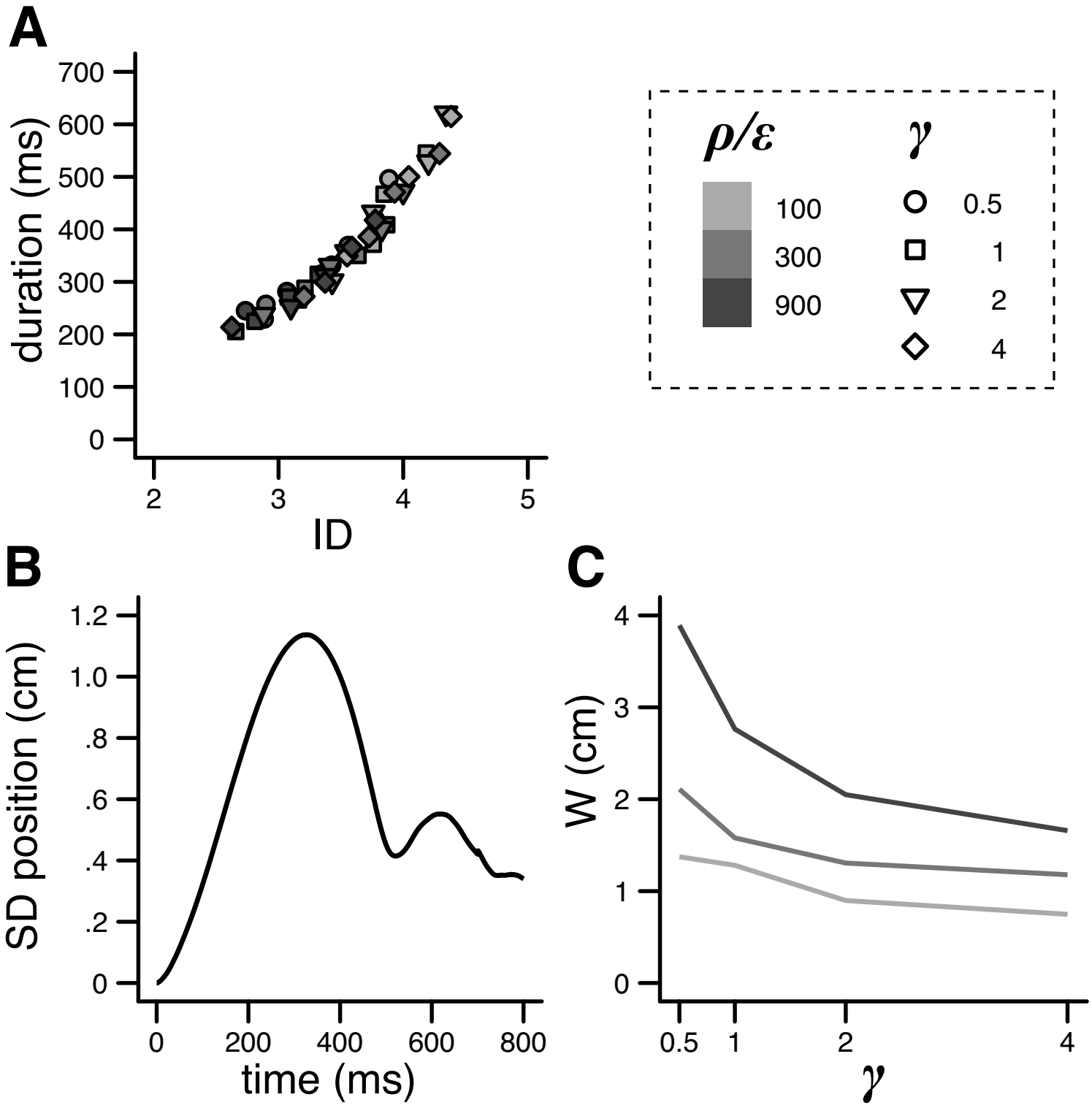

A
?
reference
test distance
B
utility
indifference
point
reference
test
distance (cm)
C
ρ/ε
tamarin
marmoset
γ
D
species
T
M
indifference point (cm)
E
large reward choice
distance (cm)


A
B
C
D
y (cm)
x (cm)
duration (ms)
distance (cm)
direction (deg)
0
5
10
15
20
25
30
90
180
270
360
100
200
300
400
500
600
−15
−10
−5


A
unperturbed
perturbed at 100ms
perturbed at 200ms
perturbed at 300ms
y (cm)
5
0
−5
0
10
20
30
x (cm)
B
speed (cm/s)
100
50
0
0
100
300
600
800
time (ms)
C
duration (ms)
750
700
650
none
100
200
300
perturbation (ms)


A
B
C
speed (cm/s)
80
0
0°
45°
90°
135°
0
500
time (ms)
y (cm)
15
10
5
0
−5
−10
−15
90°
0°
−15 −10 −5 0 5 10 15
x (cm)
1000


A
90
70
50
30
10
utility
ρ/ε decreases
.35
1.05
1.75
2.45
distance (cm)
B
90
70
50
30
10
utility
γ decreases
.35
1.05
1.75
2.45
distance (cm)
C
2.5
2.0
1.5
1.0
0.5
0.0
duration (s)
ρ/ε decreases
.35
1.05
1.75
2.45
distance (cm)
D
2.5
2.0
1.5
1.0
0.5
0.0
duration (s)
γ decreases
.35
1.05
1.75
2.45
distance (cm)


A
duration (ms)
700
600
500
400
300
200
100
0
2
3
4
5
ID
ρ/ε
γ
100
300
900
0.5
1
2
4
B
SD position (cm)
1.2
1.0
.8
.6
.4
.2
0
0
200
400
600
800
time (ms)
C
W (cm)
4
3
2
1
0
0.5
1
2
4
γ