TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Psychologie Professur für Neuroimaging

Dynamic computational models of risk and effort discounting in sequential decision making

Dario Cuevas Rivera

Dissertation

Dynamic computational models of risk and effort discounting in sequential decision making

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

dem Bereich Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

M.Sc. Dario Cuevas Rivera

geboren am 29.10.1984 in Chihuahua, Mexiko

eingereicht am 04.03.2021

Die Dissertation wurde in der Zeit von September 2013 bis März 2021 im Institut für Allgemeine Psychologie, Biopsychologie and Methoden der Psychologie in der Professur für Neuroimaging unter der Betreuung von Prof. Dr. rer. nat. Stefan Kiebel angefertigt.

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Psychologie Professur für Neuroimaging

Dynamic computational models of risk and effort discounting in sequential decision making

Dario Cuevas Rivera
Born on: 29th October 1984 in Chihuahua, Mexico

Dissertation

to achieve the academic degree

Doktor der Naturwissenschaften (Dr. rer. nat.)

Supervisor
Prof. Dr. rer. nat. Stefan Kiebel

Submitted on: 4th March 2021

List of publications used in this thesis

Chapter 2 is based on the publication:

Cuevas Rivera, D., Ott, F., Markovic, D., Strobel, A., and Kiebel, S. J. (2018). Context-Dependent Risk Aversion: A Model-Based Approach. *Frontiers in Psychology*, 9
Project conception was by DC and SK. Experimental design by DC, SK, and AS. Data gathering and programming of experiment by FO. Data analysis and modeling by DC, SK, and DM. Manuscript preparation by DC, SK, AS, FO, and DM.
This publication is not currently used in any other dissertation, nor is it intended to be used in any future dissertations.

Signature graduate student Signature supervisor

Chapter 3 is based on the publication:

Cuevas Rivera, D., Strobel, A., Goschke, T., and Kiebel, S. J. (2020). Modeling dynamic allocation of effort in a sequential task using discounting models. *Frontiers in Neuroscience*, 14
DC, SK, AS, and TG: paradigm design and manuscript preparation. DC and SK: model design. DC: experimental data measurement. DC, SK, and AS: data analysis.
This publication is not currently used in any other dissertation, nor is it intended to be used in any future dissertations.

Signature graduate student Signature supervisor

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Psychologie Professur für Neuroimaging

Abstract

Traditional discounting models in decision making research are typically applied in the specific scenario that people obtain reward after making a single decision. However, in most decision making scenarios in our natural environment, people obtain a reward only after a sequence of decisions. In these scenarios, it is still an open question whether traditional discounting models can be applied, as forward planning might affect the way risk and effort are taken into account. Another open question is whether future effortful or risky actions are considered differently from immediate ones. To address these questions, I combined computational and experimental work to test how one can model the decision making behavior of human participants in sequential decision making tasks. Specifically, I present two behavioral studies in which participants only received reward when specific goal conditions were met after a number of trials. In the first study I addressed the question of how decisions under risk can be modeled precisely by explicitly including the state-dependent context in each decision. In the second study, I established a computational model for making context-specific decisions under risk and effort. Importantly, in both the computational models developed in these two studies I explicitly included forward planning to model the way how people make a sequence of goal-directed decisions. By fitting these models to participants' choice behavior, I show the advantages of model-based data analysis for future experimental studies, as well as the usefulness of Bayesian model comparison to select the best-fitting mechanism through which future risk and efforts are taken into account when making a decision.

Contents

List of Figures

List of Tables

1 Introduction

The conscious human mind is able to cope with a never-ending barrage of small and big decisions in its everyday operations. These decisions can have consequences in time scales ranging from milliseconds (e.g. closing the eyes when something approaches them) to years (e.g. when deciding on a career path). The considerations necessary to make these decisions can be mostly intrinsic (e.g. scratching an itch) or may require taking into account large-scale systems like the public transportation in a new city or multiple players in a game.

In addition to these complications, the mind must cope with an immense number of possibilities from which to choose. Even the apparently-simple task of walking from home to the store around the corner presents the deciding mind with an almost-infinite amount of possibilities; the complete set of possible choices at any moment include all the 360° in which one could move, all the possible places in which one could change moving direction and all the ways in which one could go around obstacles.

In order to evaluate every choice available when making a decision, one must take into account the consequences of each decision, which is further complicated by consequences which are stochastic, or whose exact nature is not known to the decision maker. Some consequences are not immediate, but could have known or unknown delays associated with them.

How does the brain accomplish the impressive feat of making these decisions in an online fashion? How can one decision-making body perform in such disparate scenarios as playing chess, riding a horse, or a decision on macro-economic issues? Understanding how the human mind tackles decision making would help us better understand human behavior, with applications including but not limited to medical care (Gallagher and Updegraff, 2012), addiction treatment (Everitt and Robbins, 2005), and marketing (Nasiry and Popescu, 2012). Furthermore, it would allow us to build machines that can perform as well as humans can in decision-making tasks, with applications as simple as chatbots (e.g Serban et al., 2017) and as grandiose as space exploration (Clarke, 1968).

Risk and effort permeate every decision we encounter in real life: on the one hand, risk (and the similar concept of uncertainty) is present in every action and consideration we take. From noise in our movements and perceptions (e.g. we do not always catch the ball mid-air, even when we are paying attention to the game) to the inherent risk in any social interaction (e.g. the flat mate washes the dishes only half the times), any plan of action and any one-off decision must take risk into consideration. On the other hand, every action comes at an intrinsic cost: the effort necessary to carry it out. Cooking dinner implies the use of physical energy, which is a limited resource that needs to be regulated. Even planning our day and whether to cook dinner or eat out requires the expenditure of cognitive resources, which are thought to be finite (cf. Inzlicht and Berkman, 2015; Kurzban et al., 2013).

In what follows, the subjects of risk and effort (defined in the following sections) in decision making are described in the framework of sequential decision-making tasks. This is done across two peer-reviewed articles I published on decision making with risk and effort. In both published works, sequential decision making was used instead of the single-trial decision making paradigms that are typically used in the field to explicitly model how participants' avoidance of risk and effort can be adapted to the context in which the decision has been made, without resorting to time- or context-dependent discounting parameters.

The study of human decision making has a long history in psychology and economics. In economics, decision-making under risk (see next section) has been highly prioritized, from the earliest theories of expected utility (Bernoulli, 1954), to the formalization of the problem (Von Neumann and Morgenstern, 1944), to the initial formulation of prospect theory (Kahneman and Tversky, 1979), to the most recent versions of it (Tversky and Kahneman, 1992). In psychology, a stronger emphasis has been made on the learning of tools to make decisions, starting with the earliest studies into habitual behavior (for a review, see Anderson, 2000, chapters 2-4) and their computational implementations (e.g. Sutton and Barto, 2018; Dayan, 1993; Watkins and Dayan, 1992). In parallel to economics, psychology saw the introduction of discounting models for risk/uncertainty and temporal delay (for a review, see Green and Myerson, 2004) and, more recently, effort (Kool et al., 2010).

Many of the computational models developed through the years have been single-use models: models designed to explain behavior in a specific task (e.g. decisions under risk or delay). This includes models mentioned above, such as prospect theory and discounting models. However, additional effort has been put into developing more generic frameworks for human behavior which can be adapted to different types of decisions to be made. The earliest examples of these include dynamic programming and reinforcement learning (for a review, see Sutton and Barto, 2018). Later on, the drift diffusion model (Ratcliff and McKoon, 2008) was introduced which, while originally conceived for evidence-accumulating perceptual decision-making, has been adapted for cognitive decision making (e.g. Pedersen et al., 2017). More recently, attractor dynamics-based decision making has been proposed as an alternative to the drift difusion model (Bitzer et al., 2014). In a different vein, planning as inference, active inference and the free-energy principle (Toussaint and Storkey, 2006; Friston et al., 2014; Kaplan and Friston, 2018; Botvinick and Toussaint, 2012; Friston et al., 2015) were introduced as a possible mechanism of decision making in the brain, with wide applicability due to the generic formulation of these approaches.

Until recently, most studies on decision making under risk and effort have been done through single-trial experiments. These experiments (e.g. n-armed bandits (Averbeck, 2015)) are useful for isolating the mechanisms through which, for example, the presence of risk devalues the subjective evaluation of reward. However, most decisions in life are not made in isolation, but rather are a part of a large sequence of actions whose rewards might not come until many steps into the future. For example, the action of getting the coffee canister out of the cupboard has value only as a part of the sequence of actions that leads to a cup of coffee. The action of putting one foot forward has no value but as part of the sequential act of walking, and even walking is typically seen as part of bigger sequences (e.g. going shopping).

An important question for human decision making is whether embedding decisions in sequences of decisions (i.e. sequential decision making) requires us to adapt our understanding of the ways in which risk and effort affect our valuation of rewards which are no longer a direct consequence of a single decision. If a reward can only be obtained by completing a number of tasks, each of which comes with a probability of failure, do humans discount the reward for each one of those tasks? Or does the discounting process happen only once with an "overall" probability? Does mentally separating an effortful task into sub-tasks change how we discount the reward based on the overall effort? Can a single, fixed discount function produce changes in a person's attitude towards risk and effort throughout a task?

In this work I tackle these questions in two ways. First, in chapter 2 I present a model for sequential decision-making based on active inference (Friston et al., 2015) in which risk-aversion is modeled implicitly (i.e. without an explicit discount function), which allows the model to consider future risky decisions in the same way as immediate risky decisions are considered. Second, in chapter 3 I present a model based on single-trial discounting models for both effort and risk, with an explicit mechanism for forward-planning. With this model, future efforts and risks are modeled explicitly via two possible mechanisms each, for a total of four possible combinations of future risk/effort considerations. A novel sequential decision making task under risk and effort is also introduced, with which I was able to select the mechanism for the consideration of future effort and risk that best fit the data.

Furthermore, I focused on developing and applying generic frameworks for sequential decision making that could be applied to a number of different tasks and even combined for entirely new tasks. I show how having these models available when analyzing data allows researchers to ask novel questions that traditional data-analysis methods used in psychology cannot address. Even in complex tasks, where a single participant may not observe a significant portion of all possible scenarios during the experimental session, model-based data analysis allows for both within- and across-participant analysis, as well as obtaining trial-specific results, which is difficult –and often impossible– with traditional data-driven analyses.

With a strong focus on Bayesian methods of model comparison and fitting, I show how these models can be applied to sequential decision-making tasks and their parameters fitted to individual participants.

Before introducing the models used, it is important to formally define the two main components of the decision-making tasks to which the models were applied: risk and effort.

In both of the studies presented below, there is a strong focus on the computational models that were developed. While the computational models presented are different from each other, they have three key points in common: (1) they can be fitted to each participant separately, (2) they model the behavioural choice data of each trial, as opposed to model summary measures like the average response frequency and (3) they can be used to infer the participant's confidence on their choice, i.e. if we think of their choice as being sampled from a distribution at each trial, what was the probability of the observed choice. This last point is the focus of the first study presented below, where we argue that it can be used as a powerful data analysis tool that allows us to extrapolate choice behavior in yet-unseen contexts.

1.1 Decision making under risk

The two studies on which this dissertation is based study the concept of risk in sequential decision making. Here, I give a brief overview of the concept of risk in decision making, as well as the ways in which risk has been studied and modeled in psychology and other behavioral sciences.

Given a decision to make, risk is defined as the possibility that an action does not have the desired outcome, with a probability $p \in (0, 1)$. Examples of risky decisions, i.e. decisions in which risk exists, are easy to come by. The most obvious translation of risk into real life is in lotteries and slot machines, in which a single action (buying the lottery ticket or sliding the coin into the slot machine) has a desired consequence (winning money), but also the possibility of a second outcome (not winning money), both with a sometimes known probability. Risk, however, is found in all facets of life and can involve action sequences. For example, if running late, taking a bus (which involves a long series of single steps, e.g. walking to the bus station) gets you on time with a probability, and with X delay with a probability that depends on X. Taking a taxi (with its own sequence of actions) has the same possible outcomes but with different probabilities each.

The study of decision making under risk is of great importance as, in real life, risk permeates all the decisions that we make. From the broadest decisions (e.g. which career to pursue) to the most essential ones (e.g. how to hold a glass of water), there is always a chance that things will not go how we wanted and planned and, in many cases, alternative possibilities can be foreseen and taken into account. In the interest of understanding how humans perceive and handle risks in their decisions, risk has been studied in psychology, economics and neuroscience for decades, as introduced above.

In behavioral experiments, the typical setup for studying decision making under risk comprises a set of states in which the participant can find herself, and a small set of available actions, each of which having a set of outcomes with known probabilities. In some cases, uncertainty is also considered, i.e. outcomes to an action whose probabilities are not known to the participant and can only be inferred/approximated. A common example of this is the one- or multiple-armed bandit (e.g. Averbeck, 2015). This setup is an abstraction of the aforementioned slot machine, in which the participant is presented with one or more "arms" to pull (in analogy to old-school slot machines) in order to receive reward with a probability that is sometimes known (risk).

Such setups are used in conjunction with descriptive models that calculate a subjective value of reward, with the goal of inferring participant-specific parameters that describe how much they devalue reward based on risk. These models, often called discounting models because the reward is *discounted* (i.e. lowered) by risk, are parametric models which can be fitted to experimental data to derive parameter values that describe the risk-aversion or proneness of a participant. The best-fitting of these models is hyperbolic discounting (Ostaszewski et al., 1998), which states that a person will assign a subjective value to a reward which diminishes with risk following the hyperbola $R_{sub} = R_{obj}/(1+\kappa\gamma)$, where γ is the odds against gaining that reward, $\gamma = (1-p)/p$, where p is the probability of getting the reward. Finally, κ is the discount parameter which is fitted to a participant's choices. If $\kappa \in [0, 1)$, the participant is said to be risk prone, $\kappa \in (1, \infty)$ means the participant is risk averse, and $\kappa = 1$ means the participant is neutral, or optimal, as then the subjective value is reduced to the expected value of that action, i.e. $R_{sub} = pR_{obj}$.

It has been shown that a person's discount parameters can change over time (Mather et al., 2012; O'Brien and Hess, 2020; Green et al., 1999b, 1996). This can have two explanations: (1) the strength with which humans discount rewards based on risk changes over time and circumstances (i.e. κ depends on time), or (2) discounting itself is fixed, but the way subjective values are used to make decisions change depending on context. Note that these two alternatives are not mutually exclusive and could both be in effect.

For the two studies presented here, I chose to focus on the second mechanism, namely the adaptation of behavior to changing context. To this effect, we present two different models that exemplify how fixed discounting parameters can lead to highly dynamic behavior. In particular, in the second study, traditional discount curves are used for probability and effort discounting to explicitly fix the discounting curves to their current state of the art in the literature, while focusing on context adaptation.

In what follows I briefly introduce the published works separately. The full published texts can be found in Chapter 2 and Chapter 3.

1.2 Modeling context in sequential decision making under risk

In the first published work presented below (Chapter 2), published as Cuevas Rivera et al. (2018), we presented a behavioral model for decisions under risk and uncertainty based on active inference (Friston et al., 2015), which explicitly models the changing context for every trial and adapts its decisions. Using a sequential decision making task previously presented by Kolling et al. (2014), we used the behavioral model to model participants' choices in the task. The task is a modified k-armed bandit, in which the goal is not only to maximize reward, but also to obtain a minimum reward within eight trials (otherwise, all reward is lost). At each of these trials, only two bets are available, and the probabilities and possible payoffs are shown to the participant. At each trial, one of the bets is high-reward, high-risk, while the other is low-reward, low-risk, with the exact probabilities and rewards changing randomly from trial to trial. Participants must develop a strategy to balance reward and risk throughout the eight trials.

The model presented takes into account the accumulated points, as well as the minimum points necessary to achieve and an estimation of the possible future bets to find the strategy that balances reward and risk. The focus of this work was on the advantages that come with explicitly modeling context as part of the behavioral model, as well as modeling the preferences and beliefs of the participant in each trial. In particular, we focused on the interpolation that can be done with the models after they have been fitted to participants' data. Because many tasks in real life are complex in nature, including many possibly-hidden states of the environment, as well as actions available and changing circumstances, it is unlikely that during the course of a single experiment that aims at mimicking real life, the participant will encounter a significant portion of all the available states. This leaves the experimenter in a situation in which analyzing data on a single trial basis is difficult and oftentimes impossible.

We first showed that our trial-based analysis can reproduce the analysis performed previously by Kolling et al. (2014), which is based on a grand average across all participants and by clustering all trials into four categories. We then show that this grand-average analysis, while useful in many situations, produces results that can be misleading, as they lump together many decisions that are made in wildly differing contexts. We finalize by showing how the interpolation can lead to new insights on the behavior of participants.

1.3 Constant risk and effort aversion as dynamic observables

While not as widely studied as probability discounting, the related effect of effort discounting follows similar principles (Prévost et al., 2010; Phillips et al., 2007). Much like with risk, a reward is made less desirable, the higher the effort required to obtained is, both for physical and cognitive effort. Note that effort can sometimes be seen as its own reward, thus adding to the objective reward (cf. Wang et al., 2017). Following the steps of probability and delay discounting, different descriptive models have been presented for effort discounting

(Prévost et al., 2010; Skvortsova et al., 2014; Phillips et al., 2007; Hartmann et al., 2013; Klein-Flügge et al., 2015; Kivetz, 2003). Although effort presents extra challenges (e.g. that effort is sometimes regarding as a reward in itself (e.g. Inzlicht et al., 2018)), it is also an open question whether the parameters for effort discounting change in time and with the context, or whether they are fixed, as discussed above with risk.

In the second published article presented below (see Chapter 3), published in (Cuevas Rivera et al., 2018), we present a novel sequential decision making task involving risk and mental effort. As before, in this task participants must accumulate a minimum number of points across ten trials, to gain monetary reward. At each trial, the participant might choose a bet with a fifty percent chance of awarding a point (zero points otherwise), and the exertion of mental effort, though a number-sorting task, which always yields a point. The parameters of the task were set up such that in half the mini-blocks (collections of ten trials), participants could accumulate enough points to earn monetary reward by just choosing the bet, thus avoiding the exertion of any mental effort. This was done to determine whether people would choose to exert mental effort even when it was not absolutely necessary. This is reflected in the task in the form of a participant exerting effort during the first trials, instead of choosing the bet and waiting until the later trials to exert effort, if it became necessary.

We predicted that participants would show different strategies regarding when in a mini-block they chose to exert effort. Matching this, we found that participants could be classified as belonging to one of three groups: (1) early-effort, (2) late-effort or (3) all-effort. As their names suggest, early- and late-effort participants were those that chose to start the miniblock with effort, or end with it (when necessary), respectively. The all-effort population comprises those participants who chose effort in every trial, regardless of the contingencies of the monetary rewards.

In addition to the experimental task, we presented a behavioral model that incorporates known discounting functions and adds a forward-planning component to enable it to make sequential decisions in the task. The purpose of the model was to show that static discounting parameters, as in traditional discounting models, can create a dynamic avoidance of risk and effort. More precisely, we show that by explicitly Modeling the context in which decisions are being made, a decision making agent armed with a fixed discount function for risk and effort can make very different choices as context changes in a forced binary choice where one option involves risk and the other effort.

We used the experimental data to select the best-fitting mechanism for forward-planning, which describes how participants consider future efforts and bets to make the present decision. Finally, we showed that the fitted model parameters correlate with the dynamics of effort exertion for each participant, describing whether they belonged to the early- or late-effort groups.

2 Context-dependent risk aversion: a model-based approach

2.1 Abstract

Most research on risk aversion in behavioral science with human subjects has focused on a component of risk aversion that does not adapt itself to context. More recently, studies have explored risk aversion adaptation to changing circumstances in sequential decision-making tasks. It is an open question whether one can identify evidence, at the single subject level, for such risk aversion adaptation. We conducted a behavioral experiment on human subjects, using a sequential decision making task. We developed a model-based approach for estimating the adaptation of risk-taking behavior with single-trial resolution by modelling a subject's goals and internal representation of task contingencies. Using this model-based approach, we estimated the subject-specific adaptation of risk aversion depending on the current task context. We found striking inter-subject variations in the adaptation of risk-taking behavior. We show that these differences can be explained by differences in subjects' internal representations of task contingencies and goals. We discuss that the proposed approach can be adapted to a wide range of experimental paradigms and be used to analyze behavioral measures other than risk aversion.

2.2 Introduction

It is typically assumed that humans, as well as other animals, prefer courses of action free of risk and uncertainty; e.g., when foraging for food, easier and safer patches are preferred (Kacelnik and Bateson, 1996; Myerson et al., 2003). However, this safety-seeking behavior is highly contextual: In many situations, the course of action with the least risk or least uncertainty is not the one that can best fulfill the current goals. For example, while an animal might choose to go for small, easy prey at the beginning of a day, towards the end, if sustenance is not ensured for the night with the small prey, bigger prey must be sought, with all the risks it entails (Kacelnik and Bateson, 1996; McNamara and Houston, 1992).

Personality traits related to such risk proneness and risk aversion are well studied in humans. However, it is known that, as with animals, the preference towards risky or safe choices is not a static parameter of behavior; instead, this preference seems to change to better fit the context. While much research has been conducted on these dynamic context effects in animal behavior (Caraco et al., 1980, 1990; Cartar and Dill, 1990; Kacelnik and Bateson, 1996) and in anthropology and related fields (Winterhalder and Smith, 2000), relatively few studies seem to have explored the same themes in human subjects (Houston et al., 2014; Kolling et al., 2012; Mobbs et al., 2013).

A possible reason for this is that studying the dynamics of risk aversion in an individual is notoriously difficult (Kellen et al., 2016). This is mostly due to the fact that risk aversion must be measured, by definition, in a situation in which risk and uncertainty play a great part. In these situations, the behavior of the subject is stochastic (Rieskamp, 2008), which may complicate the analysis of the behavioral data.

To work around this difficulty, experimenters have resorted to experimental manipulations to indirectly assess or directly ask for the subject's preferences or uncertainty (Hey and Orme, 1994), which comes with its own set of pitfalls (Charness et al., 2013). Other methods are based on averaging behavior across many decisions, sometimes across many subjects and tasks, and inferring how behavior changes, on average, as a function of context, e.g. (Economides et al., 2015; Kellen et al., 2016; Kolling et al., 2014; Schwartenbeck et al., 2015; Walasek and Stewart, 2015). This approach has the downside of being blind to subject- and/or trial-specific changes on choice preference.

Here we performed a model-based analysis of the adaptation of risk aversion to context based on subject-specific behavioral responses. To do this, we combined a behavioral model formulated in the recently-developed active inference framework (Friston et al., 2015) with maximum-likelihood estimators of subject-specific parameters. The behavioral model allowed us to estimate a subject's preference for risk at every decision. Importantly, the proposed method has the advantage of not requiring multiple observations of the same context in order to estimate a subject's preferences, instead harnessing statistical power from every decision made across all trials.

As an experimental proof of principle, we applied the resulting model-based technique to a sequential decision-making task first presented in (Kolling et al., 2014). In this task, subjects must make a sequence of decisions to accumulate points towards a target. This task is well suited to study the effects of context on risk aversion, as the risk of not reaching the target varies throughout the sequence of decisions, thereby prompting subjects to adapt their choices to the current risk context.

With this approach, we find inter-subject differences in the way that context modulates risk aversion, as well as motivational and confidence-related differences in the way subjects evaluate a context and make a decision.

2.3 Methods

2.3.1 The task

Subjects performed a game-like task first introduced in (Kolling et al., 2014), in which they have a total of eight trials (decisions) to accumulate points, in what we call a mini-block. For each mini-block, a threshold is set, of which subjects are informed; if at the end of the mini-block the number of points accumulated does not exceed the threshold, all points for that mini-block are lost. The overarching goal for the subjects is to accumulate as many points as possible over all mini-blocks.

	Risky option				Safe option			
	Prob. of success	Reward	Expected value	Subj. value	Prob. of success	Reward	Expected value	Subj. value
1	0.35	265	92.75	0.243	0.9	100	90	-0.449
2	0.35	260	91	0.13	0.6	180	108	0.491
3	0.45	240	108	0.98	0.9	115	103.5	0.134
4	0.45	190	85.5	-0.149	0.6	150	90	-0.677
5	0.35	245	85.75	-0.208	0.75	145	108.75	0.215
6	0.2	350	70	0.210	0.55	145	79.75	-0.123
7	0.4	245	98	0.442	0.75	170	127.5	0.118
8	0.3	210	63	-0.165	0.9	120	108	0.329

Table 2.1: The eight action pairs offered to subjects in each mini-block, in random order. Each row represents a single action pair with one safe option and one risky option. The column 'Prob. of success' indicates the probability of getting the reward and the column. 'Reward' indicates the magnitude of a reward in points. The column 'Expected value' lists the probability of success times the reward. The column "Subj. value" shows the subjective value (see main text for the definition).

In every trial, the subjects are presented with two choices; they can either choose (i) an action that yields a small number of points with a high probability of success, or (ii) one which yields a high number of points, but with a low probability of success. Each yields zero points if it does not win. We call these two choices the 'safe choice' and the 'risky choice', respectively. Following Kolling et al. (2014), there are eight pairs of a safe and risky choices (called action pairs; see Table 2.1), and they are presented in random order in every mini-block, without repetition. The order is not known to the subjects. Subjects are informed that the outcomes of the two bets at each trial are randomly selected, independently of each other.

The maximum possible number of points in a mini-block is 2,005, which is achieved only in the unlikely case when the subject chooses and wins every risky option. The threshold to be reached, for each mini-block, is pseudo-randomly chosen from four possibilities: 595, 930, 1035, and 1105.

In this task, the variables relevant to making a decision are the trial number, the number of points accumulated so far in the mini-block, the current threshold and the presented action pair. All this information is shown to the subject on the screen during a trial (see Figure 2.1). On the top of the screen, a bar shows the subjects how many points they have accumulated through the current mini-block, as well as the threshold. We used a bar to prevent subjects from calculating exactly how many points were needed, in order to maintain uncertainty in the decisions. The two available actions are shown on the left and right of the screen; the position of the two choices (right or left side) was randomized. The probabilities of success for the actions are displayed as vertical bars, and the reward magnitudes are displayed as numbers. The trial number of the ongoing mini-block is displayed on the bottom.

Subjects made choices by pressing the X or M keys on a standard keyboard, to choose the option on the left or right, respectively. After a decision was made, feedback was displayed that informed the subjects whether the two choices had been successful or not (regardless of which choice was made). If the selected action was successful, the points were added to the top bar in white. At the beginning of the next trial, the white bar turns the same color as

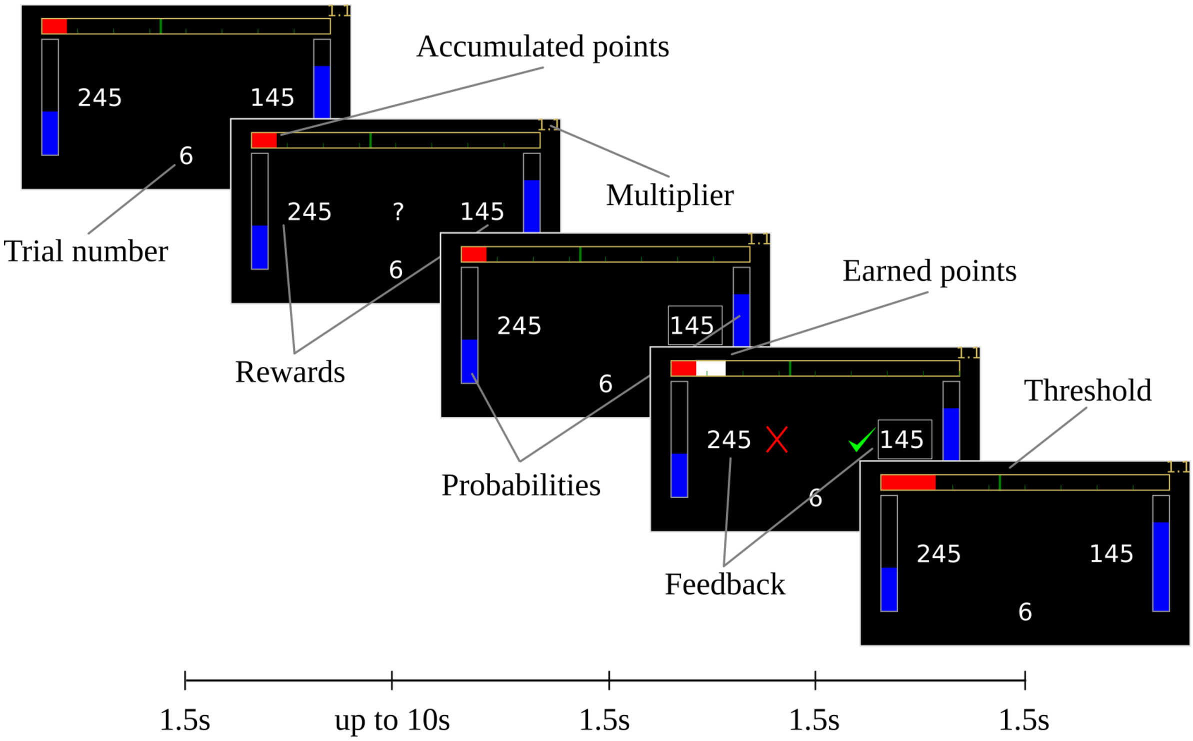

Figure 2.1: The information was shown to subjects on a computer screen. Each screen was shown for 1.5s, except for the decision screen (the second screen), which was displayed either for a maximum of 10 seconds or until the subject made a decision. If no decision was made during the 10 allotted seconds, another screen (not shown) reminds the subject to stay attentive, and the trial is repeated. After the final screen is shown for 1.5s, a new trial begins. At the end of each mini-block (of eight trials), the subject is informed about success or failure of reaching the threshold, i.e. the required number of points for this mini-block. For visual clarity, the elements are shown bigger in this figure than in the experiments.

the rest of the bar. At the end of the mini-block, subjects were informed as to the number of points gained in that mini-block. The timings of the screens can be seen in Figure 2.1.

Following Kolling et al. (2014), a mini block-specific multiplier was used, which can be seen in the top-right corner of the display. This number is set according to the current threshold, with values 1.1, 2.3, 3.3 and 4.2 (for the corresponding thresholds in ascending order). If the subject goes past the threshold of this mini-block, the accumulated points are multiplied by this multiplier. This is done to offset the difficulty of the higher thresholds and maintain the subject's motivation.

For this task, we defined the context as the combination of trial number, action pair (offered), number of points that have been earned so far, and threshold for the current mini-block. These are the variables that are relevant to making a decision and may prompt subjects to adapt their behavior in a specific context. For example, subjects may decide differently about a specific action pair when late in a mini-block with many points required to surpass the threshold, as compared to early in the mini-block. We use this definition of context throughout this paper.

The task was performed by 35 subjects, recruited from a pool of subjects at the Technische Universität of Dresden. 13 were men and 22 women, with an average age of 26 years (standard deviation 4.6). All had normal or corrected-to-normal eyesight.

The training session consisted of 4 mini-blocks, each with a different threshold. After the training session, each subject completed a total of 48 mini-blocks, 12 with each threshold, in randomized order. The session was divided into two blocks with 24 mini-blocks each, with a pause between the two blocks, totaling around 45 minutes per subject, depending on the time they took for each decision. Both the offers and their outcomes were chosen randomly before the experiments and were used with all subjects. The position of the risky option (left or right hand side of the screen) was randomly chosen for each trial.

The payout was of 10 Euros, not tied to performance. While some studies have found evidence that real vs. hypothetical payouts affect behavior in gambling tasks (Locey et al., 2011; Xu et al., 2018), it is unclear if and how adding non-hypothetical rewards would affect behavior in our task, as risky and safe choices can both be valid ways of winning a mini-block. Relatedly, whether monetary incentives alleviate or completely eliminate these biases is still under investigation, with studies attempting to eliminate the effects of these biases seeing mixed effects (Fantino et al., 2007; Locey et al., 2011). Moreover, incentivizing subjects has been found to exacerbate certain cognitive biases and hurt performance (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001).

The study was approved by the Institutional Review Board of the Technische Universität Dresden (protocol number EK 541122015) and conducted in accordance with the declaration of Helsinki. All subjects gave their written, informed consent.

2.3.2 Active inference model

Our behavioral model is based on active inference, as was described in (Friston et al., 2015). In this and the following sections, we briefly describe the Active Inference framework, as well as the generative model parameters that we used, and the fitting procedures to fit the model to each subject's data.

The active inference agent chooses actions which minimize expected free energy (i.e. maximize model evidence or minimize surprise), given the generative model and the goals of the task. The generative model is a formal description of the agent's knowledge about the existing hidden states of the environment and the existing rules that define transitions between these states. Model inversion based on Bayes' theorem allows us to formulate the agent's beliefs about the current and past states of the world, and to generate expectations about the future.

In practice, exact probabilistic inference is rarely computationally tractable, as generative models capture complex task dynamics. Hence, one often has to resort to an approximate inference scheme when defining model inversion. Active inference is based on the so-called variational approximation which allows treating posterior beliefs over specific hidden variables as conditionally independent from other factors of the hidden states space (Beal, 2003; Daunizeau et al., 2009). This method allows us to obtain closed-form algebraic equations that define the evolution of beliefs over the hidden states of the world.

In active inference, the problem of choosing an action that best fits the goals of the task is cast as an inference problem. In fact, actions are taken as another hidden state (so-called control states) of the environment, and as such benefit from the same simplification that the closed-form update equations bring.

In the following sections, we discuss the underlying hidden Markov model in which the task contingencies are represented. Full details of the mathematical derivation of the update equations can be found in (Friston et al., 2015).

Generative Model

The full generative model of active inference is built with the following Friston et al. (2015):

- A set of observations
- A set of hidden states and actions
- A generative model over observations, states and actions
- An approximate posterior probability over hidden states

In what follows, we describe these required components of the generative model in more detail.

Observations and hidden states

We take the hidden states of the environment to be two-dimensional. The first dimension is that of the accumulated points: it describes how many points in the mini-block the agent has won so far. These are taken to be the integers in the interval (0, 200), where 0 means that no points have been so far earned, and 200 is the maximum number of points possible. In the experiments, the points run from 0 to 2005; however, for the sake of computational efficiency, we divided the points by 10 and rounded up. In trial simulations, we found that this approximation made no difference in the behavior of the agent, and it allowed us to simulate the task in computation times adequate for our fitting procedures. Additionally, we did not include all 200 states (points) in the model, but rather created a cut-off point shortly after threshold; this allowed us to further reduce the computation times without affecting the results. The value chosen was 1.2 times the threshold for most of the computation, except for the (much smaller) computations in the section Section 'Risk preferences for the low- and high-STP groups', where it was set at 1.5 times the threshold to fully accommodate the consequences of all action pairs. Very few observations were made by subjects beyond the cut-off point (a maximum of 12 out of 384 for one subject) and they were removed from the pool.

The second dimension of the hidden states is the current offer; the action pairs are labeled from 1 to 8, as in Table 2.1.

We have chosen the observation likelihood, which defines dependence of observations on hidden states, as an identity matrix. This was to reflect the fact that, during the experiments, subjects receive direct observations pertaining to the number of points so far accumulated and the current offer. This means that the observations are fully informative about the state of the environment; hence the inference about the current state corresponds to matching beliefs to observations.

Generative model of states and action

The state transition probability defines the agent's belief about the evolution of the environment, both as a consequence of the agent's actions and of the passing of time.

We now describe the agent's generative model pertaining to the task at hand that consists of accumulating points throughout the trials. It is beyond the scope of this work to explain how an agent (or subject) comes to build (learn) this generative model. Here we assume that subjects have learned an accurate representation of the environment, which would correspond to setting the agents' generative model to correspond to the true generative process of the environment (i.e. the exact transition rules of the environment). This reflects the fact that the rules of the task are simple and well explained to the subjects at the beginning of the experiment.

The evolution of the environment is described in terms of transition matrices. There is one for each of the available actions. In our model, there are two actions available: risky and safe. The effect of each action depends on the current offer, which is why it is included as a dimension of the hidden states. The matrices can be represented with the following equation:

$$B_{s/r}(X_t + R_{s/r,j}, X_t) = P_{s/r,\text{succ},j} \tag{2.1}$$

$$B_{s/r}(X_t, X_t) = P_{s/r,\text{fail},j} \tag{2.2}$$

$$B_{s/r}(X, X_t) = 0, \forall X \notin \{X_t + R_{s/r,j}, X_t\} \tag{2.3}$$

Where $B_{s/r}$ is the transition matrix for the safe or risky action, respectively, X_t is the current number of points, and $P_{r/s,\text{succ/fail},j}$ and $R_{r/s,j}$ are the probabilities of success/failure and the reward of the j-th action pair, respectively. These reflect the rules of the game, in which choosing an action can either yield that action's reward, with the action's probability, or yield no reward.

Regarding the current offer (action pair), we chose a generative model which differs from the generating process of the environment. In the environment, transitions from one offer to another are randomized, but no action pair is repeated during a single mini-block. However, because keeping track of the offers already seen (and those still to come) is a very costly process, the agent instead believes that all transitions (i.e. from the current action pair to all others) are equally probable. We believe that this generative model is more likely to resemble that of human subjects.

Prior preferences over goal states

The goals of the task are stated as a distribution over the last state, after the eighth trial. The agent will make its decisions by comparing its predictions regarding states to be visited in the future by making a set of decisions, and the prior distribution over the last state.

In our case, this distribution describes the relative desirability of outcomes of a mini-block, in terms of the accumulated points. A score is assigned to each of these hidden states of the environment; the higher this score, the more the agent will seek to be in this state by the end of the mini-block. We assumed that subjects learned this distribution through both task instructions and experience during training.

Given the task instruction one could reason that the best functional form for the preferences is the one with zeros everywhere below threshold and some positive number above threshold (see Figure 2.2A). However, it is important to note that the prior preferences over the goal state do not necessarily reflect the rules of the game, but rather establishes the behavioral strategy that the agent believes is the best for winning. Therefore, we expect that, through experience, subjects build different beliefs as to what the best way to win is. Because of this, we allowed for subject-specific goal distributions.

For our fitting procedures, we used three distinct shapes for the goals: Gaussian, sigmoid, exponential. By changing their parameters, we were able to generate a large family of shapes that are consistent with the task instructions. In particular, the sigmoid family recreates the task instructions to the letter, such that any state below threshold has no value, and a ramping up holds for those above threshold. Examples of these shapes can be seen in Figure 2.2B-C.

The sigmoid family has two parameters: slope and center. The center determines where the transition from zero to one is centered, and the slope parameter is the slope at the center. These follow the equation:

$$f_{\text{sig}}(x) = \frac{1}{1 + e^{-m(x - x_0)}} \tag{2.4}$$

Where m and x_0 are the slope and the center, respectively. A very high m and a x_0 at threshold reproduces the step function as in Figure 2.2A.

The exponential family has only one parameter, the coefficient of the exponent. This family follows the equation:

$$f_{\text{exp}} = e^{kx} \tag{2.5}$$

where κ is a free parameter, For the remainder of this work, we will refer to the parameter κ of the exponential goal shape as sensitivity to points (STP), in analogy to the sensitivity to delay/probability in the discounting literature (Basile and Toplak, 2015). This family describes a 'ramping-up', which does not incorporate the existence of a threshold. This shape represents a simple heuristic, where the agent assigns exponentially larger amounts of preference to higher number of points.

Finally, the Gaussian family has two parameters: mean and standard deviation. They are given by the equation:

$$f_g(x) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.6}$$

where μ and σ are the mean and standard deviation, respectively, and A is the normalization constant. This family reflects a compromise between the general rule of 'more points

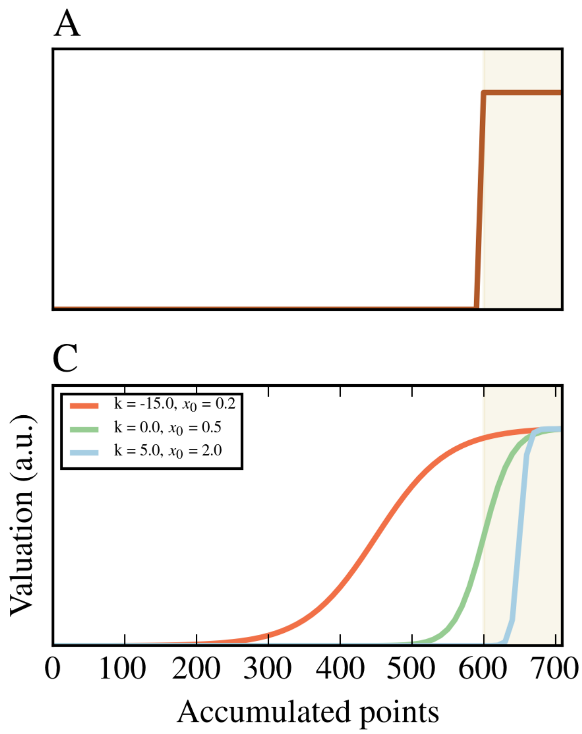

Figure 2.2: Goals are expressed as a valuation of each of the relevant hidden states, which in this case are the accumulated points after the 8th trial. These valuations are presented here as shapes, following the equations described in the Methods section. The shade areas are those points above threshold. For visualization, only points from 0 to 700 are plotted, with a threshold of 595. (A) An example of a shape that follows the task instructions to the letter. All points below threshold have a value of 0, while all above are valued the same. (B) Gaussian, (C) sigmoid and (D) exponential families of shapes. For each family, three examples are plotted (shown in different colors), with their corresponding parameters shown in the legend. The values of μ and κ are relative to the threshold. For example, $\mu = -150$ represents a mean located at $\hat{x} - \mu = 445$. In (D), STP stands for sensitivity to points, which is the parameter of the exponential family (see the section Section 'Subject-specific parameter values' above).

are better' and a strong preference to states close to threshold (depending on the mean of the distribution).

Note that, while most of the shapes considered do not have a strong threshold that the agent must surpass (with the exception of the sigmoid, when the slope is high), they do not conflict with the rules of the task; by maximizing points (which can be said of any monotonously-increasing shape), a mini-block can be won. Additionally, certain parameter ranges (for example, higher values for the exponential) do include the idea of soft thresholds, i.e. that the agent has a strong preference for ending above threshold.

Posterior over actions (update equations)

To calculate the posterior probability distribution over the available actions at each trial, $p(a_t|s_t, m)$, our model makes use of the following equation:

$$\log P(a_t|s_t) = F(s_t, C) + \log(\beta_{a_t}) \tag{2.7}$$

where is a subject-specific choice bias, which does not depend on the context and the function $F(s_t, C)$ compares the projected future states to the agent's goals C; see (Friston et al., 2015) for details.

We introduced this choice bias parameter because we observed a marked average preference of most subjects for the safe choice across the entire data set (see Section 'Standard analysis of behavioral data' below). Choice bias is a prior preference for or against the risky option, regardless of context and, more specifically, of the current offer. That is, has two components, β_{risky} and β_{safe}. The values referred to in the main text are those of β_{risky}. This parameter complements the goal parameters by adding a component that is non-contextual on top of the contextual goal parameters. For an overview of the model parameters, see Figure 2.3A.

2.3.3 Fitting the model

Our model has free parameters which we fitted to every subject independently by doing a grid-search over the relevant part of the parameter space. These parameters are of two categories, affecting the agent in different ways, see Figure 2.3A.

In the first category, the first parameter affects how extreme the probability distributions from which actions are sampled can be. We call this parameter maximum decision modifier (MDM) and it corresponds to the parameter in (Friston et al., 2015). The effect of the MDM is that, for a low value, the distributions approach 50/50, regardless of context, while for large value, they approach either 0/100 or 100/0, depending on the context. This can be interpreted as controlling how certain the agent can be about its decisions; its value, either optimal (performance-wise) or inferred (fitted to a subject) is both task- and subject-dependent. The second parameter in this category is choice bias, which is a non-contextual number added to (or subtracted from) the preference for risk.

The second category pertains to the shape of the goal distribution. These are the three families discussed above, each with its own parameters.

For each subject, we performed a grid-search over the parameter space and calculated, for each set of values for the parameters, the data likelihood of the model. Through this procedure, we created a likelihood map for each subject, which represents a multi-dimensional probability distribution over parameter values.

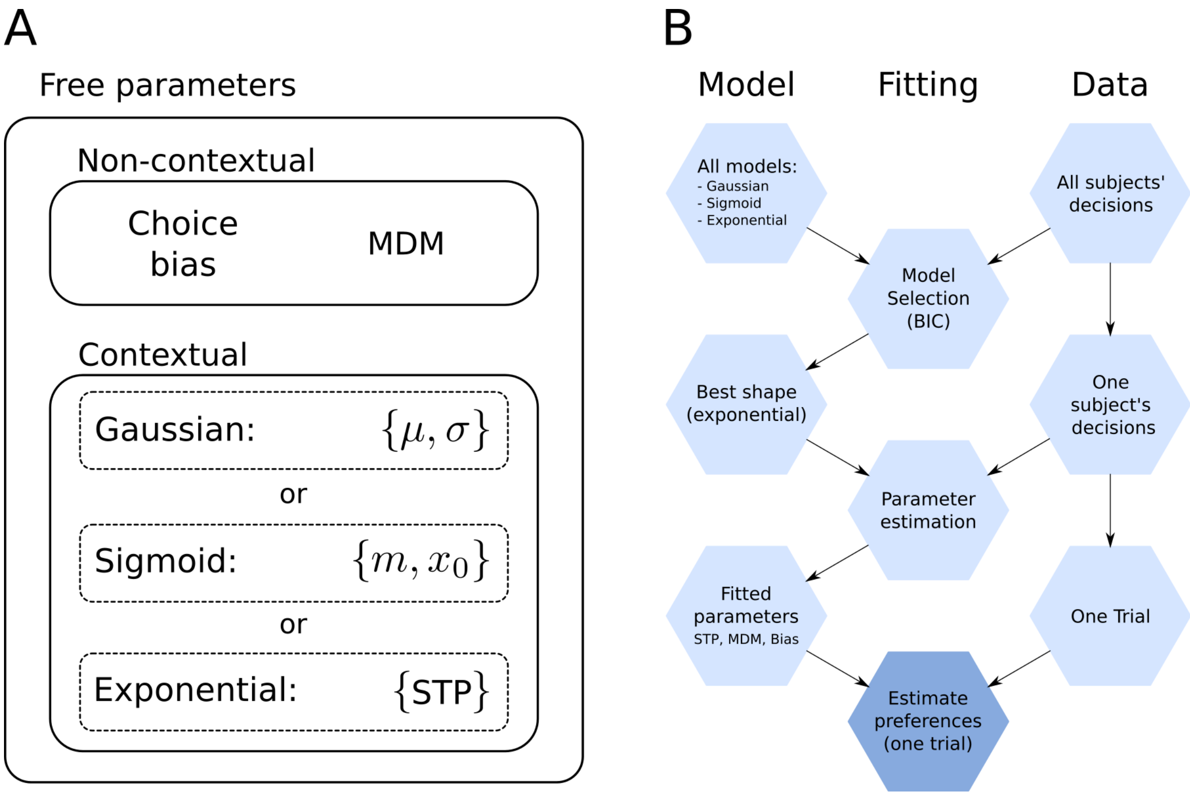

Figure 2.3: (A) Model parameters. The goal shape (e.g. exponential) is shared by all subjects, while the rest (Subject-specific, free parameters) are fitted to each subject. Non-contextual parameters are those that affect all decisions equally (the choice bias and the maximum decision multiplier, MDM). Contextual parameters are those whose effect on decisions changes with context, which are those related to the shape of the goal. The number of parameters depends on the shape; the Gaussian has two (μ, σ), the sigmoid two (m, x_0) and the exponential one (STP) (see Section 'Prior preferences over goal states'). In total, the model with the exponential has four free parameters and with Gaussian and sigmoid five. (B) Diagram for the proposed method to estimate subjects' preferences. The process starts at the top, with the collection of all models and the experimental data for all subjects. Nodes on the left-hand column are models, which become more refined as the process moves downwards. The same is true for the Fitting and Data columns (center and right-hand, respectively). The final step, "Estimate preferences", is repeated for each trial, while the rest are performed only once.

The parameters we searched are as follows. For the Gaussian family, μ from $\hat{x} - 15$ to $\hat{x} + 15$, σ from 1 to 15, where $\hat{x}$ is the threshold, both with increments of 1. For the sigmoid family, we chose centers x_0 between $\hat{x} - 15$ and $\hat{x} + 15$ with increments of 1, and slope m from 0.1 to 3 with increments of 0.1. We chose to set the values of μ and x_0 to be centered about threshold in order to unify the parameter values across all conditions and simplify calculations; since the interpretation of the fitted agents is made in terms of their shapes and not their parameter values, this choice has no effect on the results. For the exponential family, the parameter STP was taken from 0.5 to 10, in increments of 0.1. Smaller increments in these parameters caused no discernible differences in the posteriors over actions.

The parameter MDM was searched in the range 0.1 to 5 in increments of 0.1, and from there to 60 in increments of 5. The range of MDM is segmented in this way because for higher values (¿5), small increments create no discernible differences in the posterior distributions over actions.

Finally, the choice bias parameter was searched between 0.1 and 2 in increments of 0.1. Maximum likelihoods rapidly drop after 1.2, and by 2 are already many orders of magnitude lower.

2.3.4 Model comparison

We made use of the Bayesian information criterion (BIC) (Schwarz, 1978) to evaluate each model and compare between them, using the guidelines in (Kass and Raftery, 1995).

In Figure 2.3B we show a diagram of the entire procedure comprising model comparison, parameter fitting and preference estimation (which will be discussed below).

2.4 Results

The influence of the environment, as well as the internal state of an agent, e.g. a foraging bee, on the decision-making process under varying risk has been extensively studied in the fields of behavioral ecology (for animals) and behavioral anthropology (for humans) (Winterhalder and Smith, 2000). In these studies, animal or human subjects are placed in environments, or presented with tasks, in which a sequence of decisions must be made in order to reach an overarching goal. For each decision, two or more options are presented to the subject, differing in how risky these options are, i.e. how likely it is that they will yield no reward, and how much reward, e.g. food or money, they may yield. Here we define risk as the probability that a chosen action does not yield any reward, when these probabilities are known to the decision-making agent (as opposed to uncertainty, where the contingencies are unknown). Because this definition of risk (and its usage in risky/safe choices) refers to the differences on the probabilities of success for the two choices in a trial, "relative risk" would be a more adequate name; however, since it is the only definition of risk used in this study, we have dropped the adjective "relative" from the name for simplicity.

For this work, to exemplarily showcase this approach, we made use of one such task, introduced in (Kolling et al., 2014), in which the variations in the context, i.e. the states of the environment and of the agent relevant to the decision-making process, allowed us to study the effects of context on the choices of human subjects. More specifically, it allowed us to study the adaptation of risk aversion to changing context, as has been done with animals (Caraco et al., 1980, 1990; Cartar and Dill, 1990; Kacelnik and Bateson, 1996).

This section is divided into three parts: firstly, we show, using standard analysis tools, the behavioral results. Secondly, we present a computational model for the task, as well as the results from fitting this model to behavioral data. Finally, we show that our approach can be used to track the change in subjects' risk aversion throughout the experiment.

2.4.1 Standard analysis of behavioral data

We first performed the same data analysis as was performed in (Kolling et al., 2014) on our data set of 35 subjects.

In Figure 2.4A, we show the average probability of choosing the risky action as a function of ΔV, the difference in subjective values of the risky and safe choices. To reproduce the findings in (Kolling et al., 2014), we defined the subjective value of a choice as:

$$V_{\text{risky/safe}} = \tilde{p}_{\text{risky/safe}}(\text{win}|\text{offer}) + \tilde{r}_{\text{risky/safe}}(\text{offer}) \tag{2.8}$$

where and are the probability of success and the reward for the current offer, where offers are normalized by the mean and standard deviation across all offers. The tildes in and indicate that they have been normalized: for each offer, we subtracted the mean across all offers and divided by the standard deviation (see Table 2.1). Note that the term "subjective" refers to the fact that it differs from the objective definition of expected value (probability multiplied by reward); Kolling et al. (2014) opted for this definition as they found that it better fitted subject's choices. Note that, to show the expected value in the conventional sense, Table 2.1 lists the standard expected values for each offer pair, based on a multiplication of probability and reward size.

As already found in (Kolling et al., 2014), it can be seen that, when averaged across all subjects and all mini-blocks, the probability of choosing the risky offer increases monotonously with the difference in subjective values. Note that ΔV does not offer a full description of a context, for it does not take into account the potential pressure created by the upcoming end of the mini-block or the necessity to go above threshold in the task (Kolling et al., 2014).

Kolling et al. (2014) defined the variable 'risk pressure' as a more detailed (but still incomplete; see below) description of the context of the present task. Risk pressure is defined as the average number of points the subject would need to earn in each of the remaining trials in the mini-block in order to surpass the threshold:

$$\varGamma_t = \begin{cases} \frac{\hat{x}-x_t}{T-t} & \text{if} x_t < \hat{x} \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

where t is the current trial number, is the threshold, xt the accumulated number of points at t, and T the total number of trials. They found evidence, using a regression analysis, that its value is a predictor of subjects' choices. Risk pressure has values in the interval [0,]: (i) zero when points are above-threshold, (ii) and when no points have been earned and only one trial is left. In the present task, risk pressure values above 350 represent a context in which the subject cannot possibly win, as the highest possible offer is 350 (see Table 2.1).

We replicated the behavioral results in (Kolling et al., 2014) using two fixed-effects generalized linear models (GLMs); both assume the same regressor values for all subjects. The first model had four regressors, namely: a constant term (choice bias), ΔV, trial number and risk pressure. The results can be seen in Figure 2.4B. In our case, probably due to the larger number of subjects, all regressors are significant (p¡0.001); however, to exactly reproduce

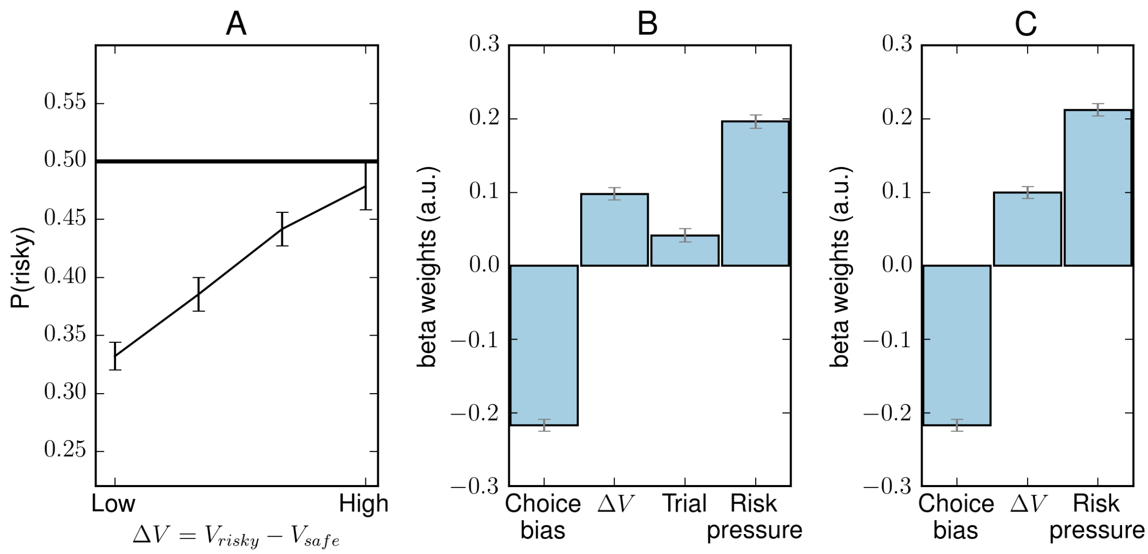

Figure 2.4: Behavioral results. These results reproduce the results reported in (Kolling et al., 2014). (A) Probability of choosing the risky choice, P(risky), as a function of ΔV, the difference between the values of the risky choice and the safe choice, averaged across all subjects. (B) Fixed-effects logistic regression with decisions pooled from all subjects as the dependent variable (1 for risky, 0 for safe) and four regressors: a constant term (choice bias), the difference in values for the two choices (as in (A)), trial number and risk pressure. (C) GLM regression as in (B), but without the trial number as regressor. Error bars are SEM. All regressors in (B) and (C) are significant (p¡0.001).

the previous results we created a second GLM with three regressors: constant term, ΔV and risk pressure. The resulting regression can be seen in Figure 2.4C.

2.4.2 Adaptability of risk aversion

In this section, we use a simple method to analyze the adaptation of risk pressure, where many trials are binned together and averaged to calculate the subject's preferences for risk aversion for this set of trials. This is an obvious extension of the analysis performed in (Kolling et al., 2014) for ΔV (also see Figure 2.4A). We show the limitations of such a method and show, in subsequent sections, how the model-based approach we present in this paper sidesteps these difficulties.

When trying to estimate subjects' preferences from a single decision in a stochastic environment, that single decision is not necessarily representative of the subject's preference. This is especially true when the preference is not too strong, which we found to be the case for this task. In such a case, to estimate the small size of the preference, one would require many exposures to this decision in the same context. However, this is suboptimal due to potentially confounding memory effects Bornstein et al. (2017); Mather et al. (2003): the subject may remember having made a specific context-dependent decision before, which will influence any subsequent repetitions.

In order to study the context dependence of risk aversion, a simple analysis method is to make use of risk pressure as a description for the context: We binned risk pressure to obtain enough decisions to calculate a risk pressure-dependent risk aversion value for each bin.

Any such reduction of dimensionality will have the disadvantage of possibly mapping two very distinct contexts onto the same risk pressure value. For example, a risk pressure of 300 on the last trial, when an option to win 350 points has been offered to the subject, presents a context with 20% probability of going above threshold (see Table 2.1) if the subject takes the risky offer, and 0% if the subject takes the safe offer. The same risk pressure near the beginning of a mini-block would present a game in which winning is almost impossible, and choice becomes inconsequential. This downside is inevitable in order to obtain reliable estimates on the mean risk aversion when using averaging methods.

As can be seen in Figure 2.5A, high values of risk-pressure above 350 were observed much more rarely than lower values. This is a direct consequence of the task design: if values of risk pressure above 350 were commonly observed, subjects might lose motivation due to the high difficulty of the task.

We calculated risk aversion as a function of risk pressure, by binning the contexts for each subject according to values of risk pressure. For each bin, the average risk aversion can be calculated by averaging across all decisions made in the bin. In Figure 2.5B, the frequency of choosing the risky choice for each bin is shown for three representative subjects, along with the 95% confidence intervals. The results for the remaining subjects can be seen in the supporting information (sup. fig. 1). We found that for 25 out of 35 subjects, the null hypothesis that risk aversion is constant for all values of risk pressure cannot be rejected (one-way ANOVA, p¿0.05).

In what follows, we introduce a novel method for inferring the risk aversion on a trial-by-trial basis.

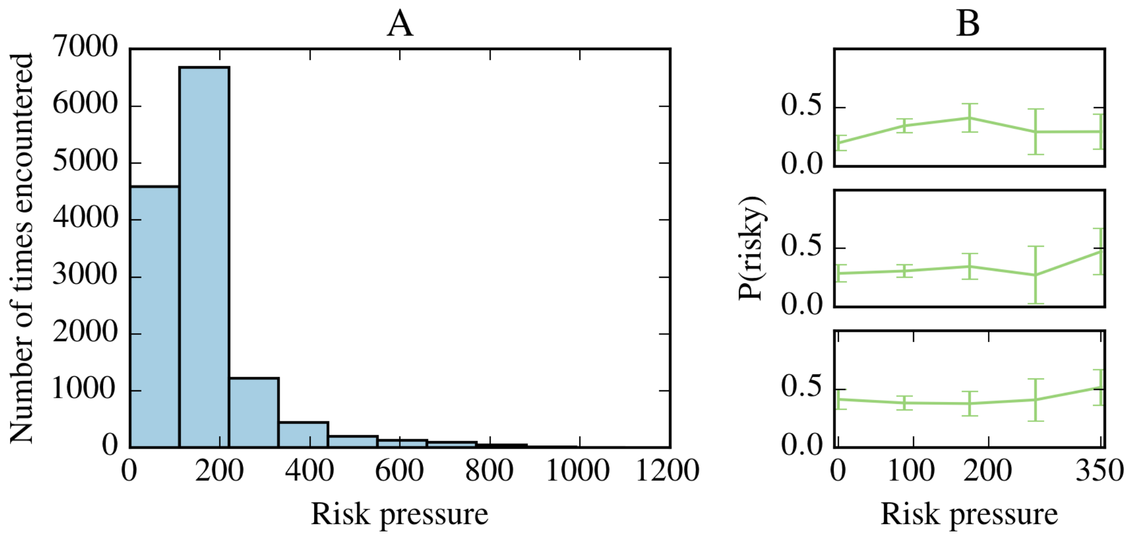

Figure 2.5: Risk pressure and risk aversion. (A) Histogram of the values of risk pressure encountered by all subjects. (B) Adaptive risk aversion for three representative subjects where we plot, against risk pressure, the frequency that the subject chooses the risky option, P(risky), calculated by averaging over all choices for each subject, binned according to the context in which the choice was made. Only risk-pressure values between 0 and 350 are shown as only few contexts were observed in higher values of risk-pressure by all subjects –see (A).

2.4.3 Model-based approach

In this section we describe a novel method for estimating the subject's preference (i.e. risk aversion) for every context they observed. This method takes advantage of all decisions made by the subject to calculate the subject's preference in any given context. We introduce a model-based approach using active inference, with free parameters that we fitted to each subject using all the decisions made by the subject (see Section 'Methods'). This so-called agent (the fitted model) can then be used to estimate, for any given context, the probability distribution over actions that is most consistent with the subject's entire set of choices.

This process is equivalent to finding the underlying mechanism with which subjects make their decisions (which is assumed not to change throughout the experiment). Once this mechanism has been found, it can be used to calculate the preferences (risky vs. safe) that the subject had for any trial.

We will show that these fitted agents can be used to study in greater detail the changes of risk aversion in human subjects.

2.4.4 Model parameters and fitting

We fitted three different models which differed in their parameterization of the goal shape; see Section 'Methods'. For each of these three models, we fit the model's parameters to the choices made by the subjects throughout all the mini-blocks. We fitted a total of four parameters to subjects' choices: (i) the maximum decision multiplier (MDM), (ii) up to two parameters that control the shape of the goal distribution, and (iii) a choice bias parameter.

For more details on these parameters, see Section 'Methods'.

The MDM and choice bias parameters differ from the goal shape parameters in that their effect on choices does not depend directly on the context. For this reason, their effects will be discussed mostly separately from those of the goal shape parameters.

2.4.5 Modeling inter-subject differences

We first hypothesized that previously observed inter-subject differences in risk aversion (Choi et al., 2007; Shead and Hodgins, 2009) can be modelled for this task here as differences in the internal representation of the task contingencies. For example, an agent for whom finishing below threshold is unacceptable (Figure 2.2A) will tend to display, if necessary, riskier behavior than one who is simply trying to increase the number of points (blue and brown line in Figure 2.2).

To address this hypothesis, we tested for inter-subject differences in the model parameters that best fit the behavior of each subject. To do this, we compared two models: With the first, we inferred a single, best-fitting parameter set for the data of all subjects, i.e., same parameter values for all subjects. For the second model, we inferred subject-specific parameters. We found that the BIC difference between these two models strongly favored the subject-specific model (ΔBIC ¿ 100), as per the guidelines in (Kass and Raftery, 1995).

Given that there was strong evidence for inter-subject variability in how subjects represented the goal of the task, we next performed a model comparison between the three goal shapes, to determine whether one family was significantly better at explaining subjects' behavior than the other two.

We found strong evidence that the exponential shape is better at explaining subjects' behavior than the other two families, as evidenced by ΔBIC ¿ 100, and therefore all the following results were obtained with this exponential shape family.

Finally, we tested whether there was also evidence for condition-specific differences in goals, i.e. different goal shapes for the four different thresholds, for each subject. However, we found no significant advantage in separating data into conditions; therefore, all results that follow were obtained with the subject-specific model but with the same parameters across conditions.

We also found that the introduction of the subject-specific MDM significantly improved model fit (ΔBIC ¿ 100). Adding the choice bias parameter further increased model fit (ΔBIC ¿ 100).

To summarize, we found very strong evidence for the model with inter-subject differences, with the exponential family being the best model for representing subjects' goal shapes. This model was used for the results presented in what follows.

2.4.6 Subject-specific parameter values

The MDM values we obtained for different subjects range from 0.1, which makes most decisions close to a 50/50 decision, to values as high as 20, which creates preferences close to being deterministic. In Figure 2.6A we present a histogram of the values of MDM for all subjects, where the average MDM value is 6.01 (standard deviation of 5.70). The number of bins and their positions were determined using the SciPy Jones et al. (2001) implementation of the Freedman-Diaconis estimator (Freedman and Diaconis, 1981), assuming equal-width bins.

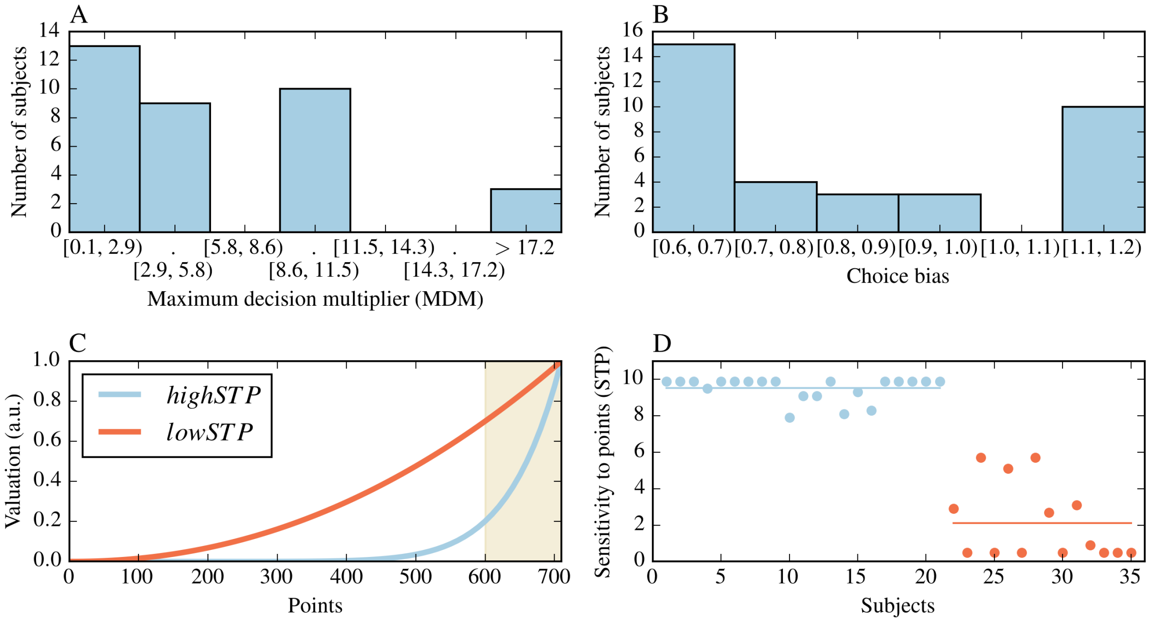

Figure 2.6: (A) Histogram for the best-fit values of the maximum decision multiplier (MDM) for all subjects. (B) Histogram of the best-fit choice bias parameter values, where the value of 1 stands for unbiased choice behavior and values smaller than one for a bias for the safe choice. (C) The shapes elicited by the centers of each of two identified clusters, as given by the cluster center's sensitivity to points (STP) value. The shaded area indicates above-threshold values (depicted here with a threshold of 595). (D) Scatter plot of the STP parameter for all subjects. The colors indicate cluster membership, as in (C). Subjects were sorted by cluster membership. The horizontal lines represent the values of STP which were determined as the cluster centers in (C).

The low values of the MDM for many subjects reflect a degree of indecisiveness, with which decisions are mostly driven by the choice bias parameter; this seems to be at least partially due to the fact that the task rarely presents contexts in which one action is clearly better than the other. Those subjects with a high MDM tend to have many strong preferences throughout a mini-block; these strongly determined decisions are the least affected by the choice bias.

All best-fit choice bias parameter values were found between 0.6, which favors safe choices, to 1.2, which favors risky choices (a value of 1 has no effect), and a mean of 0.8. For most subjects (26 out of 35), we found a best-fit choice bias parameter smaller than one, which indicates an increased preference for the safe choice.

We also found inter-subject differences in the shape parameters, reflected in subject-specific values for the STP. To summarize these results, we clustered all subjects based on their fitted STP using the k-means algorithm (Goutte et al., 1999; Lloyd, 1982). We found that the best number of clusters is 2, and the clusters resulting from this procedure can be classified as low- or high-STP; see Figure 2.6C-D. Subjects belonging to the low-STP group implement a simple heuristic of accumulating more points, regardless of threshold; those in the high-STP group are best described by representing rather sharp goal shapes, i.e. by giving relatively low importance to points below threshold. For a full list of the inferred parameters, see the supporting information (sup. table 1).

Finally, we calculated a Pearson cross-correlation matrix for the fitted parameters and found no significant correlations between them.

2.4.7 Recovering subjects' preferences

With the estimated parameters for each subject, we can build a decision-making agent (henceforth, a fitted agent) which makes decisions the most consistently with its corresponding subject. Thus, for each subject we have a fitted agent.

Using these subject-specific fitted agents, we recovered subjects' preferences via the posterior distributions over actions that the subject-specific fitted agents calculate for every context. The key advantage of this procedure is that it eliminates the necessity of having many trials measured experimentally in the same context (or the same variable of interest, which is related to context, e.g. risk pressure), when investigating subjects' trial-by-trial preferences and their dynamics. In other words, the model-based approach is like a 'mathematical microscope' (Moran et al., 2013) which enables the experimenter to replace actual (here binary) choice observations of a single subject, for a specific context, by inferred posterior probabilities lying continuously between 0 and 1.

2.4.8 Adaptation of risk aversion

In this section we show the behavior of the fitted agent in a similar manner as that used for Figure 2.5B, discuss the shortcomings of this approach and, in the next section, show how our model-based approach can be used to sidestep these shortcomings.

We first looked at the subject-specific general risk aversion, i.e. how likely subjects are to pick the safe option throughout all contexts they observed. Averaged across all the trials and subjects, we observed a probability of choosing the risky option of 0.41, with a standard deviation of 0.07 across subjects, before taking into account the choice bias parameter. Despite the differences in parameter values, we found no significant differences between subjects (one-way ANOVA, p¿0.05).

Figure 2.7 (left), where we show the recovered preferences for three representative subjects (the same as in Figure 2.5B), for the set of all the contexts that they observed. Note that the spread of the dots is not measurement noise but the inferred preferences for differences offers. While different trends can be seen for the three representative subjects (e.g. subject A's preference for risk increased for the first 100 units of risk pressure, while subject B's preference dipped in the 50-150 range), this representation suffers from similar problems as those discussed in section 'Adaptability of risk aversion'.

Relatedly, the sharp vertical swings of the averages in Figure 2.7 (left-hand side, solid lines) are due only to the fact that the observations made by subjects do not sample all values of risk pressure evenly; the calculation of the preferences (vertical positions of the dots) itself entails no stochasticity, as the model is deterministic. These variations make extracting any significant information from these plots all but impossible. However, our model-based approach allows us to overcome this difficulty by being able to predict how such a subject would behave in the "missing" contexts, uncovering the differences in behavior between subjects A and B in greater detail; See the next section.

For validation purposes, we compared the estimates obtained with our method and with the binning method (see Figure 2.5B). Since most of the observed contexts (for all subjects) are for low values of risk pressure (see Figure 2.5A), and to allow for a better comparison between the methods, we binned the decisions in risk pressure values between 0 and 200 (with bins at 0, 50, 100 and 150). The results are shown in the right-hand column of Figure 2.7, alongside the average obtained with our method (the same as the left-hand column), using the same bins as the binning method, i.e. we binned the preferences estimated by our method and calculated their average for each bin. For clarity, in the remainder of this work, we use the name "model-based binning" to refer to the results presented in the right-hand column of Figure 2.7, i.e. to binning the model-based risk aversion to calculate averages (blue line). For all three subjects both methods agree in their estimate of how often, in a specific risk pressure range, these three subjects will choose the risky choice. There are small differences, especially for subjects A and B, which we will discuss below.

2.4.9 Risk preferences for the low- and high-STP groups

In this section, we show that the jaggedness of the average risk preference for each value of risk pressure (Figure 2.7, left-hand column) is mostly caused by the biases introduced by the set of contexts observed by the subjects. We also show how one can use the fitted agents to "see through" these biases, to gain further insight into the adaptation of risk aversion to context for the two groups of subjects.

For the results in this section, we used the agents that were fitted to subjects A and B from Figure 2.7. To facilitate a direct comparison between the risk aversion adaptations of these two subjects, we set their MDM to 20 and their choice bias to 1. Because the effects of these two parameters are not context-dependent, the principled findings in this section are not affected by using the fitted values while the visual clarity of the figures is increased. Note that for both subjects, MDM values are similar in their effects (10 and 5, respectively) and choice biases are identical (0.6). We used these two modified agents to extrapolate these subjects' preferences to contexts that were not observed during the experiment.

As shown above, the model-based approach allows us to have a trial-specific estimation

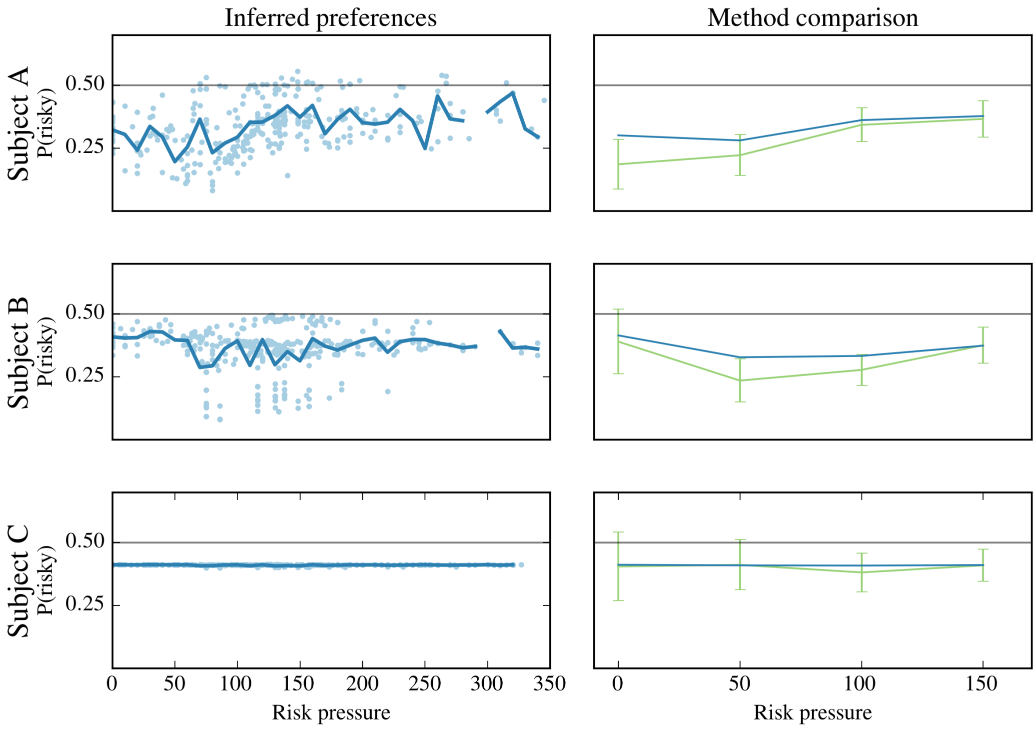

Figure 2.7: Probability of choosing risky. The three representative subjects are labeled A, B and C, and each row corresponds to one subject. These subjects are the same as those used in Figure 2.5B. Left-hand column: plot for each subject, labeled 'Inferred preferences', represents the subject's preference for the risky option. For each data point (context of being offered, under as specific value of risk pressure, a specific option pair, see Table 2.1) the subject's preference for the risky option (y-axis) is plotted against risk pressure (x-axis). For each risk-pressure value, the average preference for the risky option is calculated (solid line). Not all risk pressures have been sampled by the subjects during the experiment as can be seen from the broken solid line for subjects A and B. Right-hand column: to validate the model-based approach, we show the binning method shown in Figure 2.5 (green) and the preferences estimated by the model-based approach, averaged on the same bins as the binning method. Note that the definition of bins deviate from Figure 2.5 as only risk pressure values up to 150 are shown in these plots; this is done for a better comparison with the model-based approach (see main text).

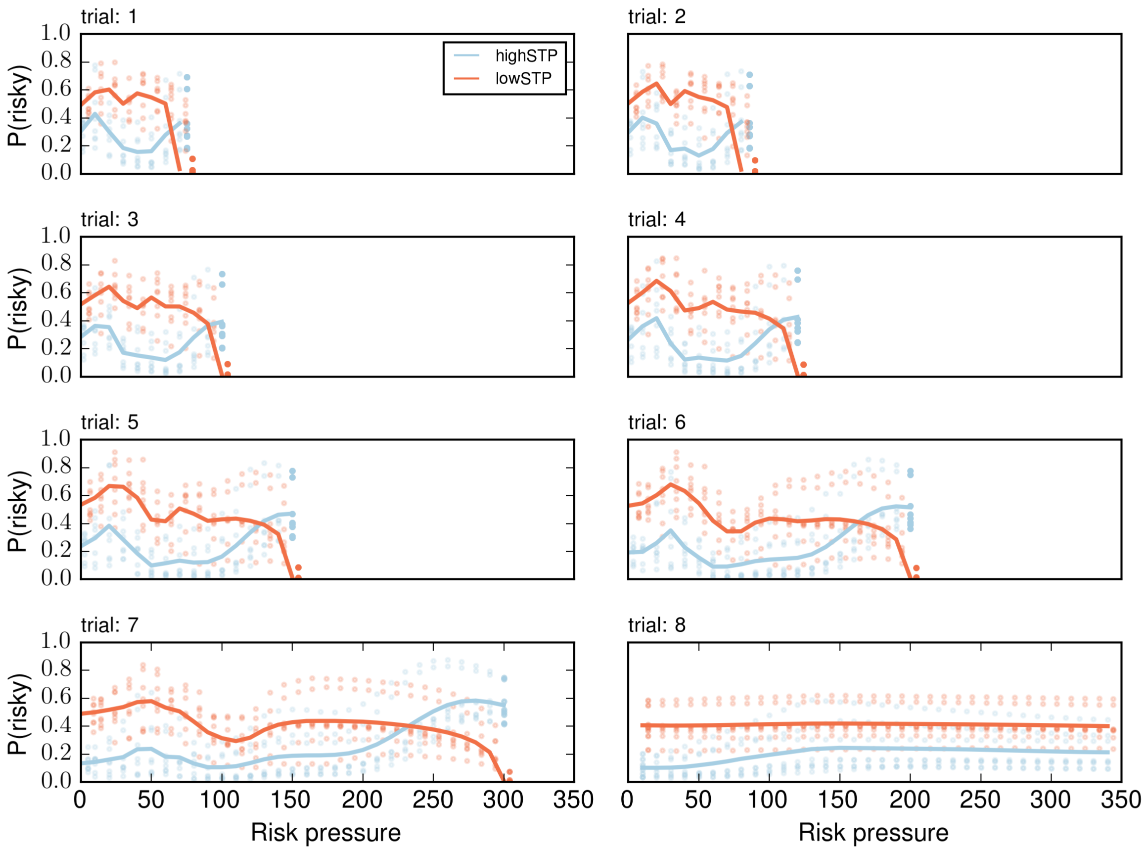

Figure 2.8: Each plot is for a different trial number, shown above each plot. Agent A, a representative of the high-STP group, is shown in blue. Agent B, a representative of the low-STP group, is shown in orange. We exposed both agents to all possible contexts in the task. Each single dot indicates the probability of choosing the risky choice for a single context. Orange dots were offset by 0.2 units to the right for visual clarity; averages (solid lines) were not offset. The solid line represents averages that were obtained for each value of risk pressure across all eight offers. For every trial, there is a maximum possible risk pressure (threshold divided by remaining trials); blank spaces on each plot are values of risk pressure beyond that maximum.

of subjects' risk aversion. This enables us to separate trials not just by risk pressure as in Figure 2.7, but also by trial number, essentially creating a two-dimensional projection of the four-dimensional context (trial number, offered action pair, number of points that have been earned so far, and threshold). Such a separation, which can be seen in Figure 2.8 for agents A and B, revealed stereotyped trends in the adaptation of risk aversion to risk pressure as mini-blocks advanced.

Importantly, for each trial and risk pressure value, the vertical spread of the points is due to the available action pairs (eight of them), and not due to measurement errors, as there is no stochasticity in these estimated preferences.

A key feature of this presentation is the stereotyped curves, across trials, for average risk aversion adaptation that differ from trial to trial only in their length of spread across risk pressure values. This difference in length is caused by the fact that it is impossible to observe high risk pressures at the beginning of a mini-block, while on later trials the risk pressure may become quite large if the subject is missing a lot of points to reach the target. For any trial number, the high-STP subject type has a higher risk aversion than the low-STP subject type for the lowest values of risk pressure. For the largest values possible in the trial, however, this relationship is inverted, the low-STP subject type having now higher risk aversion than the other. This trend holds for most trials and slowly disappears towards the end of each mini-block. As can be seen in Figure 2.7, this phenomenon is lost, for both model-based and standard binning, when averaging across all trials and using the contexts observed by subjects.

To avoid biasing the estimates of Figure 2.8 with the specific set of contexts that any subject observed, we created a set of contexts which contains exactly one context for every combination of risk pressure, trial number and offered action pair. To make use of this set of contexts, we take advantage of the fact that our agents allow us to extrapolate how the corresponding subject would react to any context, even if it was not observed during the experimental session. For simplicity, we fixed the threshold to 595 for all contexts as we did not find evidence for differences in subject parameterizations for different conditions (see above).

This set of observations is unbiased in the sense that all possible observations are encountered exactly once. The importance of this can be seen in Figure 2.8: if for any value of risk pressure the action pairs with the highest probability of choosing the risky option had not been encountered by the subject, the average for this value of risk pressure (solid line) would be much lower. Conversely, if any one action pair had been encountered multiple times, the average would be skewed towards the probability of choosing the risky option for that action pair. The model-based approach is not susceptible to such biases. Critically, this bias is outside experimental control, as it depends on the subjects' choices during the experiment, which determine what risk pressure values are experienced for each action pair and trial number. Clearly, such a bias may impact the results and interpretation when using the binning method; see Section 'Discussion'.

Figure 2.8 also reveals a trend not visible in more coarse-grained descriptions such as Figure 2.7: as mini-blocks progress, the average risk aversion across all values of risk pressure becomes flatter and more similar between the two subject types. This culminates in the last trial of the mini-block, where both average lines are very similar to each other and close to a flat line around medium values of risk pressure. We believe this to be a consequence of the smaller depth of future planning that a subject has to go through at each trial. In the last trial of the mini-block, when no future planning is necessary, the probability of choosing the risky choice peaks in very different places of risk pressure, for different action pairs, creating the flat average seen in the last trial (since this is the average across all action pairs). To make this clearer, in Figure 2.9, we plotted the probability of choosing the risky choice for four different offer pairs in the last trial. This plot contains the same data as the last panel (trial 8) of Figure 2.8, but different types of lines (e.g. dashed) are used instead of individual dots to aid in differentiating between action pairs, and the other four offers were removed to avoid visual clutter.

Figure 2.9 shows that, even when the difference in average risk aversion (averaged across all action pairs, for every risk pressure value) between the two agents is not large (as seen in Figure 2.8 for the 8th trial), the probability of choosing the risky option of a specific action pair can differ greatly between the two subject types. In general, for the high-STP subject type the probability of choosing the risky option for an action pair has steeper curves than for the low-STP subject type. This is partially due to the fact that the high-STP subject type is much more inclined to take the risky choice for any one given offer at risk pressure values in

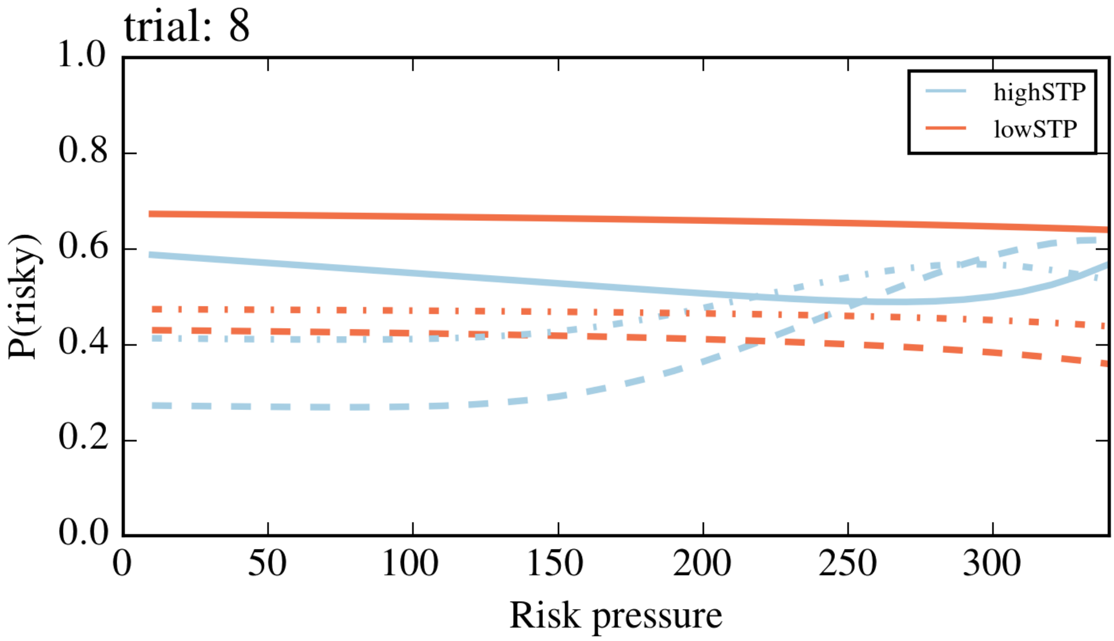

Figure 2.9: Probability of choosing the risky option for the first four offers (action pairs) in Table 2.1, during the last trial, for both subject types (orange for high-STP, blue for low-STP). The four offers are plotted, each with a different type of line, (in the order of Table 2.1): solid, dashed, dash-dotted and dotted. Lines of the same type with two different colors represent the same offer for the two different subject types. The last four offers of Table 2.1 were removed for visual clarity. For example, the dashed lines represent the offer (safe: 180, risky: 260), where the blue dashed line shows that the high STP subject type chooses more often the risky choice once the risk pressure is above 180, i.e. there are more than 180 points missing. In contrast, the low STP subject type does not change much its choice behavior when offered this specific offer on the last trial.

which the safe choice is not good enough to succeed. This is especially true if the probability of success for the safe offer is not much higher than the risky choice, e.g. for the 2nd offer (dashed line in Figure 2.9, safe offer: 180 points with probability 0.6, risky offer: 260 points with probability of 0.35; see Table 2.1).

Looking at the averages shown in Figure 2.8, it may appear as if the average differences in behavior (solid lines) between the two subject types can be explained by a potential difference in choice bias. This would shift the entire solid line up or down, if the choice bias were bigger or smaller than 1, respectively. However, it is evident from Figure 2.9 that this is not the case when looking in more detail at single offers: given the different shapes between the blue and orange curves, shifting any curve up or down (by adding a choice bias different from 1) would not make them similar across subject types.

Although we focused here on only two subject types to show principled subject-specific differences, note that these two subject types are representatives for the high-STP and the low-STP group. As we have shown in Figure 2.8 and Figure 2.9, the model-based approach enables a precise mapping between subject-specific goals and corresponding action preferences at the single-trial level for a concrete context experienced in a single trial. Here, we have used this mapping to infer from sequences of choices the internal goal function at the single-subject level.

2.5 Discussion

To reveal adaptation of risk-taking behavior to changing context, we developed a model-based approach for inferring the risk aversion with subject- and trial-specific resolution. This allowed us to use the set of binary choices made by subjects to infer the preference that each subject had when making those decisions. The method uses the computational framework of active inference and is based on a decision-making model with free parameters that can be fitted to an individual subject's decisions. To show how this works in practice, we used a sequential decision-making task, in which subjects aim to achieve a goal over multiple trials. This task was designed to elicit changes in risk-taking behavior as the context changes in which the subject must make a decision.

We demonstrated how this model-based approach can be used to analyze behavior at a subject- and trial-specific resolution, which is generally not possible with standard binning methods due to limitations in the amount of data that is typically collected. Using clustering, we found that subjects can be divided into two groups, according to a key model parameter that describes the relative value that subjects give to increasing gains. We found that these two groups have different risk-taking behavior adaptation to changing context.

2.5.1 Risk aversion adaptation

Traditionally, behavioral studies on risk aversion have focused on its fixed, context independent component. Economic theories of decision making under risk, e.g. prospect theory (Kahneman and Tversky, 1979), as well as psychological experiments on probability discounting (Green et al., 1999a), have focused on characterizing human choices in fixed, often hypothetical, contexts, and risk aversion is assumed not to change throughout the experiment.

Conversely, the dynamic, context-adapting aspect of risk aversion has seen much study in ecology, where it is established that risk aversion is not only a function of the given options, but also of the current needs and their urgency (McNamara and Houston, 1992). From ecology, risk-sensitive foraging theory (McNamara and Houston, 1992) emerged as a normative account of a foraging animal in an ever-evolving environment, which takes into account the changing needs of the foraging animal. The jump to the study of risk-taking behavior to human subjects has been made in the field of human behavioral ecology, where the tenets of theories such as risk-sensitive foraging theory have been applied to human evolution; for an extensive review, see (Winterhalder and Smith, 2000).

Despite this, descriptive accounts of risk aversion as an adapting variable of human behavior have been more limited. Examples of studies that have focused on the dynamic, context-adapting side of risk aversion in human subjects are (Kolling et al., 2014), and (Schweighofer et al., 2006), the latter in the similar field of delay discounting.

The approach in (Kolling et al., 2014) was a direct, data-driven measurement of risk aversion, which relied on binning enough decisions together to obtain reliable statistics, from which a preference for risk can be calculated. Such binning of distinct decisions is necessary in tasks with a large set of possible contexts since it becomes impossible to sample once, let alone enough times for a reliable estimation of a mean, any sizeable portion of the whole space. For example, in the task used by (Kolling et al., 2014), the number of different contexts is in the order of tens of thousands, while the number of trials per subject was only 384.

In this work, we presented a model-based approach which is not affected by the size of the set of contexts and with which there is no need for a proxy variable nor binning decisions, since estimates on risk aversion can be obtained by fitting a decision-making model to the entire data set of each subject. This approach assumes that the experimentally-measured choices are sufficient to infer about an underlying choice mechanism with which one can extrapolate the remaining contexts not seen by the subject in the experiment. As the underlying choice mechanism we used the so-called active inference framework (Friston et al., 2015). The overall procedure resulted in an increased resolution, at the single-trial level, for the study of the adaptation of risk aversion to changing context. As a beneficial side effect, the model-based approach also avoids known statistical artifacts of binning approaches, e.g. hiding existing effects in the data (Ashby et al., 1994; Cohen et al., 2008; Estes and Maddox, 2005; Siegler, 1987).

2.5.2 Subject classification and differences in behavior

Using the fitted model parameters, we found that subjects can be classified into two groups (see Section 'Model parameters and fitting'). We found differences both when looking at subjects' preferences averaged across all trials, for which the differences are shown to be only evident in the low risk-pressure ranges (see Figure 2.7), and when looking at the more detailed per-trial analysis (see Figure 2.8 and Figure 2.9).

While it is difficult to discern differences in Figure 2.7 due to the averaging over trials (see Section 'Adaptability of risk aversion'), Figure 2.8 and Figure 2.9 reveal the clearest differences. We found that there are clear differences between the choice behavior between the high-STP and low-STP subject type; for the high-STP subject type, risk aversion decreases as risk pressure increases, while for the low-STP subject type, the highest values of risk pressure see an increase in preference for risk.

The behavior of a high-STP subject type is to try and go above threshold. When more points are needed to reach the threshold (high risk pressure), more risk is taken. When few points are required (low risk pressure), risk is unnecessary and therefore avoided. We found that a low-STP group subject type shows a much weaker awareness of the threshold. It is an open question what exactly low-STP subjects try to achieve. A simple explanation may be that they were less motivated, or less goal-directed than the high-STP subject types to try and succeed in this specific task. Note that the model-based approach inferred these differences without an obvious reliance on subjects' performance: we did not find a significant performance difference between the two groups. We inferred the differences in subjects' internal representation of the task goal just by fitting the pattern of choices to the underlying active inference model. The finding that there was not a difference in performance between the two groups is possibly a consequence of the stochastic nature of the task; in future work, it may be useful to focus on less stochastic tasks to confirm the difference between the low- and high-STP groups also in terms of performance or other measures. We believe it is a strength of the proposed method that we can model the behavior of these subjects who do not necessarily follow the instructions to the letter but nevertheless perform well.

The lack of threshold awareness is noticeable when comparing the curves on the last (eighth) trial in Figure 2.8 and Figure 2.9. For the high-STP subject type, the probability of choosing the risky option for each offered action pair peaks at different values of risk pressure, depending on the action pair, which reveals threshold-awareness. The low-STP subject type instead shows little adaptation, at this last trial, of risk aversion to risk pressure.

The last trial also offers some insight into a feature of the model-based approach, as applied to this experimental task. Intuition suggests that the maximum probability of choosing the risky option should be very high in the last trial when the reward offered by the safe option is insufficient. However, the underlying heuristic of "the more the better" is present even for the high-STP subject type: This can be seen in Figure 2.6C where the goal shape of the high STP-group (blue solid line) has non-zero values even for a number of points below threshold. These non-zero values increase the usefulness of the safe choice even in cases where it is not enough to go above threshold, decreasing the preference for the better risky choice. This, coupled with the low probability of success with the risky choices, effects the lack of extreme preferences for the risky choice in the last trial seen in Figure 2.8 and Figure 2.9. Whether or not subjects' choices show the same behavior is rather difficult to determine because such cases occur rarely (around 1% of trials in our data set). Additionally, given the low number of occurrences, even a single safe choice made by a subject in these situations (e.g. by not paying attention or by miscalculation) throws off any estimate that is made based on these choices alone, or by the whole set of choices (as with our approach). Moreover, subjects had only a rather coarse-grained indicator of the number of missing points by a plotted bar as shown in Figure 2.1, so there may be cases where subjects were not entirely clear about whether a safe choice would be sufficient to reach the goal, or not. In fact, in the 1% of trials where a subject may obtain the target only by choosing the risky option, subjects chose the risky option only 75% of the times, which could explain the results seen in Figure 2.8 and Figure 2.9. It could prove interesting to explicitly model this observation noise in future studies in order to determine whether the inferred goal shapes better conform to the task's rules.

Starting from these initial results, the model-based approach can be used to generate predictions. For example, one could ask (i) how specific subjects will respond in specific but yet unseen contexts and whether this can be generalized and translated, using inferred parameters, to different experiments, and (ii) what specific contexts are those, for a specific experiment, that will show the most obvious differences between two groups of subject or, more generally, along a specific trait dimension. One can also use Bayesian model comparison to test alternative or extended versions of the current model. For example, the low percentage of risky choices observed in extreme contexts on the 8th trial might be caused by observation noise: subjects do not know exactly how many points they need to go above threshold and might therefore miscalculate the usefulness of the safe option. As a first approximation, the current model does not account for such observational noise; rather it calculates with absolute precision whether a given reward would suffice. Because of this, one way for the model to account for these decisions (safe choices in extreme cases) was to "soften" the shape of the goals, assuming that subjects found values below threshold acceptable, if not optimal, which gives a higher preference (likelihood) to the exponential family of goals over the other two.

2.5.3 Describing a context with a proxy variable

Reducing a complex context into a one-dimensional proxy variable (such as or risk pressure) can bring with it loss of information. In the specific case of the task used here, as in many other tasks used in psychology, the context, i.e. the set of variables necessary to make a decision, is multi-dimensional. In many tasks, e.g. the urn task (FitzGerald et al., 2015), the set of all contexts is small enough that a proxy variable might not be necessary, or a natural proxy variable, such as trial number itself, can be used, as e.g. in (Schweighofer et al., 2006). However, in other tasks as the one used here, in which the number of contexts is usually in the thousands, it is difficult to obtain veridical estimates on context-dependent risk aversion without using proxy variables that bin together dissimilar decisions.

With our model-based approach we do not need to use such a proxy variable for context; instead, the context is evaluated by the model by using all the information available to make a decision. This approach additionally revealed the effects of using a proxy variable. This can be seen in Figure 2.7 where, for each value of risk pressure, data points at different heights can be seen. These represent trials with different trial numbers, offered action pair, number of points and/or threshold, which nonetheless might have the same risk pressure. In standard binning approaches, these would be binned together to obtain summary statistics.

As a first attempt to validate the results from our model-based approach using standard binning methods, we compared the overall preference for the risky offer, averaged across all subjects and trials. A finer comparison was made in Figure 2.7, where we showed that binning the results using model-based binning yields curves of summary statistics that are very similar to those obtained with standard methods.

It is this similarity between the two methods that hints at important information being lost when binning many decisions: with the results from the binning method we could not conclude that any adaptation of risk aversion to risk pressure happens as we could not reject the null hypothesis of no risk adaptation for 25 out of 35 subjects. In contrast, our model-based approach reveals differences between stereotyped risk aversion adaptation curves that are clearly visible when looked at in a trial-specific fashion (Figure 2.8). Additionally, our method provided a single, subject-specific parameter value (i.e. STP, sensitivity to points) to explain the rather complex adaptation of risk aversion.

2.5.4 Risk aversion

In this work we showed that differences in risk-taking behavior can be partially explained by differences in the way that subjects set goals for themselves. These internal goals are related to the goals of the task, as they were explained to the subjects, but are not necessarily the same as determined by the instructions.

The internal representation of goals sets not only the desired end states of a task, but also the valuations of these end states relative to each other (see Figure 2.6C). It is these relative values that can explain the inter-subject differences observed in the adaptation of risk aversion. For example, a subject with an internal goal of maximizing the number of points achieved at all costs could be less risk-prone that a subject who only wants to go above-threshold, due to the expected value of the safe choice is larger than that of the risky choice (see Figure 2.7). The effects of this internal representation of goals on risk aversion is highly contextual: in a given context, if two subjects have different internal goals, one subject might show a stronger preference for the risky offer than the other, while for another context, the opposite would be true. This is what differentiates this adaptive account of risk aversion from a trait-like account. Examples of this can be seen in Figure 2.9B, where differently-colored lines of the same action pair (e.g. dotted lines) cross each other.

To complement this adaptive account, we introduced the additional parameter of choice bias, which accounts for an overall preference for safe choices (risk aversion) or risky choices (risk proneness) by using prior probabilities for actions. In this work, we implemented a subject-specific, overall prior preference for (or against) the risky choice for any action pair. This choice bias is not informed about context or about the relative values and probabilities of success of the given choices. We found this simple, fixed bias to be an important parameter as evidenced by very strong evidence for it when comparing models. In this sense, the model presented in this work incorporates an adaptive risk aversion adaptation, which tailors responses to changing contexts, and a trait-like, non-adaptive risk aversion, which biases all responses towards or away from risk taking behavior. Similarly, in future work, additional biases, e.g. choice-supportive bias or the effect of previous successes with an offer, can be added to the generative model in a straightforward way, using model selection to identify those bias parameters that provide for better models of adaptive risk aversion.

3 Modeling dynamic allocation of effort in a sequential task using discounting models

3.1 Abstract

Most rewards in our lives require effort to obtain them. It is known that effort is seen by humans as carrying an intrinsic disutility which devalues the obtainable reward. Established models for effort discounting account for this by using participant-specific discounting parameters inferred from experiments. These parameters offer only a static glance into the bigger picture of effort exertion. The mechanism underlying the dynamic changes in a participant's willingness to exert effort is still unclear and an active topic of research. Here, we modeled dynamic effort exertion as a consequence of effort- and probability-discounting mechanisms during goal reaching, sequential behavior. To do this, we developed a novel sequential decision-making task in which participants made binary choices to reach a minimum number of points. Importantly, the time points and circumstances of effort allocation were decided by participants according to their own preferences and not imposed directly by the task. Using the computational model to analyze participants' choices, we show that the dynamics of effort exertion arise from a combination of changing task needs and forward planning. In other words, the interplay between a participant's inferred discounting parameters is sufficient to explain the dynamic allocation of effort during goal reaching. Using formal model comparison, we also inferred the forward-planning strategy used by participants. The model allowed us to characterize a participant's effort exertion in terms of only a few parameters. Moreover, the model can be adapted to a number of tasks used in establishing the neural underpinnings of forward-planning behavior and meta-control, allowing for the characterization of behavior in terms of model parameters.

3.2 Introduction

It has been known for long that physical effort appears to bear an inherent cost both in humans and other animals (Hull, 1943; Walton et al., 2006). Although the nature of cognitive effort remains elusive (Shenhav et al., 2017), the role of mental effort has been studied more recently in the same vein (Kool et al., 2010; Apps et al., 2015; Pessiglione et al., 2018; Schmidt et al., 2012), as well as its neural underpinnings, e.g., (Radulescu et al., 2015). Generally, effort seems to carry a disutility that diminishes the value of reward an action entails, a phenomenon known as effort discounting (Westbrook et al., 2013; Botvinick et al., 2009).

In psychology and economics, much effort has been put into establishing so-called effort discount functions, i.e., parameterized functions that describe how the subjective value of a reward diminishes as a specific amount of effort is required to obtain it. As with delay and probability discounting, several parametric shapes of the effort discounting function have been suggested: hyperbolic (Prévost et al., 2010), inspired by delay and probability discounting; linear (Skvortsova et al., 2014); bilinear (Phillips et al., 2007); parabolic (Hartmann et al., 2013); and sigmoidal (Klein-Flügge et al., 2015). Additionally, a framework based on prospect theory conceptualizes effort discounting as a shift of the status-quo (Kivetz, 2003). See also (Talmi and Pine, 2012; Białaszek et al., 2017; Klein-Flügge et al., 2015) for comparisons between these different models.

While these studies established a mathematical description of how required effort affects the valuation of a reward, the experiments were typically constrained to the particular case where the decision to invest effort to obtain reward must be made immediately. However, in most cases of goal-directed behavior in daily life, the reward is not obtainable immediately but must be pursued over an extended time period. This means that in typical effort discounting experiments one cannot address the question of when people will invest effort to obtain a reward that remains obtainable over an extended period of time. For example, an employee may be given a deadline of two weeks to complete an assignment that takes one day. The question for this employee on every day until assignment completion is whether she should invest the effort today or wait until later (Steel and König, 2006). This question is outside the domain of typical effort discounting experiments because there is no 'wait until later' option. Some individuals would probably do the assignment early because there may be an unforeseen situation that prevents them from finishing later. Others would prefer to wait and intend to do the assignment late, e.g., just before the deadline runs out, because perhaps it turns out that the assignment is no longer required. Clearly, all possible courses of actions (do the effort early or late) have their advantages and disadvantages and put individuals into a decision dilemma. We believe that this dilemma is central to the meta-control question of how effort discounts potential reward because the dilemma emerges typically when one is pursuing goals that cannot be obtained now but only after some extended time (Goschke, 2014).

In order to induce this dilemma, it is necessary to put participants in a situation where forward planning and future contingencies are important, as opposed to the single-trial experiments traditionally used to elicit discounting. By forward planning, we mean that to make a decision one has to plan several time steps into the future to predict the consequences of possible courses of actions (Dolan and Dayan, 2013). For example, the employee may on day one simulate through in her mind several alternatives of when to do the assignment, select one of these alternatives and execute the first action of this alternative. The question is how one can model decision making in this dilemma by combining forward planning over several trials and previously established effort discounting models for a single trial.

To address this question, we developed a sequential decision making task that captures the effort-investment decision dilemma described above. In each trial of a trial sequence, participants were given the choice to exert effort right away to improve their chances of obtaining a reward at the end of the trial sequence, or wait and not invest effort to see how the situation evolves, so that eventually the need for effort might disappear, however at the price of lowering the chances of reward. We found that the proposed computational model was able to explain different time points at which different participants invested effort. Using formal model comparison, we inferred the forward-planning strategy used by participants during the task. We also show that the inferred effort- and probability-discounting parameters provided for an easily interpretable explanation of the early versus late effort allocation effect observed in the choice data.

In summary, we present a computational-experimental approach, in the form of a novel experimental task and a sequential decision-making model, that enables future studies into the effects of pursuing long-term goals based on moment-by-moment decisions about effort investment in human participants.

3.3 Methods

Participants were recruited from a pool of potential participants organized by the Technische Universität Dresden that includes students as well as individuals from the general population. Of $N = 60$ participants taking part in the experiment, five had to be excluded based on their poor performance during an initial training period (see below). This left $N = 55$ participants (18 female, with an average age of $M = 26.0$, $SD = 10.8$) for our analyses.

Participants went through two different experimental tasks which, together with introduction and training, took an average of 1.5 hours. The two experimental tasks were a single-task effort/probability discounting paradigm and the novel sequential task. In this work, we report only the analysis of the sequential task data that was performed before the single-trial task. For this reason, we describe here only the sequential task.

Payoff was a basic reimbursement of 9 Euros for participating, plus a performance-based bonus of up to 5 Euros for the sequential task. Some participants traded the basic reimbursement for course credit. On average, participants who did not trade the basic reimbursement for course credit earned around 14 Euros for the whole experiment.

The study was approved by the Institutional Review Board of the Technische Universität Dresden (protocol number EK 541122015) and conducted in accordance with the declaration of Helsinki. All participants gave their written, informed consent.

3.3.1 Sequential task

In this task, participants were instructed to accumulate points over the course of a mini-block (a trial sequence) of ten trials, with the objective of surpassing a point threshold at the end of a mini-block (displayed as an empty bar to fill with points). To do this, they had to, at every trial, choose between a mentally effortful and a probabilistic option.

If the participant chose the effortful option, she must complete a number-sorting task, in which a set of numbers was shown on screen with five digits each that can differ in any of the digits (see Figure 3.1A). The participant had to sort the set of numbers in ascending order by sequential mouse clicks on the displayed numbers within a fixed time period, the length of which was determined during training (see below). If the participant correctly sorted the numbers, a point was gained for that trial, which was shown on the bar at the bottom. No point was gained if the numbers were not sorted correctly.

If the participant chose the probabilistic option, she had to complete a number-sorting task as well, but all numbers had a single digit, rendering the task practically cognitively effortless.

If the numbers were correctly sorted, participants had a 50% chance of earning a point (and 50% of earning none), of which they were informed during the instructions. The probabilistic option corresponded to waiting until a later trial to exert effort, if it ever became necessary. The probability associated with the probabilistic option was included to create mini-blocks in which the participant could win without having to exert any effort by choosing this option at every trial and being "lucky" with the outcomes. We included the single-digit sorting trial to equalize the physical effort that comes from using the mouse to click on the numbers.

The time allotted to a participant for sorting the numbers was adapted to each participant during training such that their performance on the number-sorting task (with five digits) is around 90% (see Section 'Procedure') to equalize the required effort across all participants. To avoid time becoming a confound, this participant-specific time was the same for both the effortful and the probabilistic option, determined for each participant during training. By doing this, all trials lasted exactly the same time for each participant.

At each trial, the current number of points was displayed as a bar shown on the bottom of the screen (see Figure 3.1B) during the cue and decision phase (see Figure 3.1C). In order to fill the bar in the mini-block, five points were necessary. If during a mini-block the bar was filled, 20 Euro cents were added to the participant's final reward. Otherwise, they gained no reward for the mini-block. Each participant went through 25 mini-blocks. Monetary reward was contingent on winning mini-blocks (as opposed to simply maximizing points) to give special significance to winning a mini-block and to implicitly dissuade participants from focusing on getting the maximum number of points by always choosing the effortful option.

Each trial of the sequential task was divided into three phases: (1) the cue and decision phase (Figure 3.1B), in which participants had to choose between the two options using the keyboard ("c" for the option shown on the left, "m" for the option shown on the right). The left/right position of the two options (probabilistic and effortful) on the screen was randomized every trial. This phase lasted until the participant made the decision, but no longer than three seconds; (2) the sorting phase, in which participants had to carry out the selected task. This phase lasted between four and ten seconds, depending on the participant's performance during training (see below); and (3) the feedback phase, in which participants were told whether they correctly completed the task or not. This phase lasted half a second. Figure 3.1C shows a diagram of the trial timing, including all the screens observed by participants as well as the timings of each phase of the trials.

Importantly, the number of points required to win a mini-block was only half of the number of trials in the mini-block. This, combined with the 50% chance of getting a point with the probabilistic option, had the effect that, by just choosing the probabilistic option, the participant could win on average half the mini-blocks in the experiment. Additionally, because the difficulty of the effortful task was set such that expected performance is close to 100%, the participant was almost guaranteed to win every mini-block, regardless of the strategy chosen, as long she was willing to invest the effort associated with the effortful option when it became necessary, i.e., when she would otherwise have risked not having enough points at the end of the mini-block.

3.3.2 Procedure

The experimental session began with instructions shown on the screen. No instructions were given by the experimenter. Then, the participant went through an introduction to the number-sorting task with the intention of getting them acquainted with how the mouse is used to sort the numbers. During this familiarization period, participants completed twelve

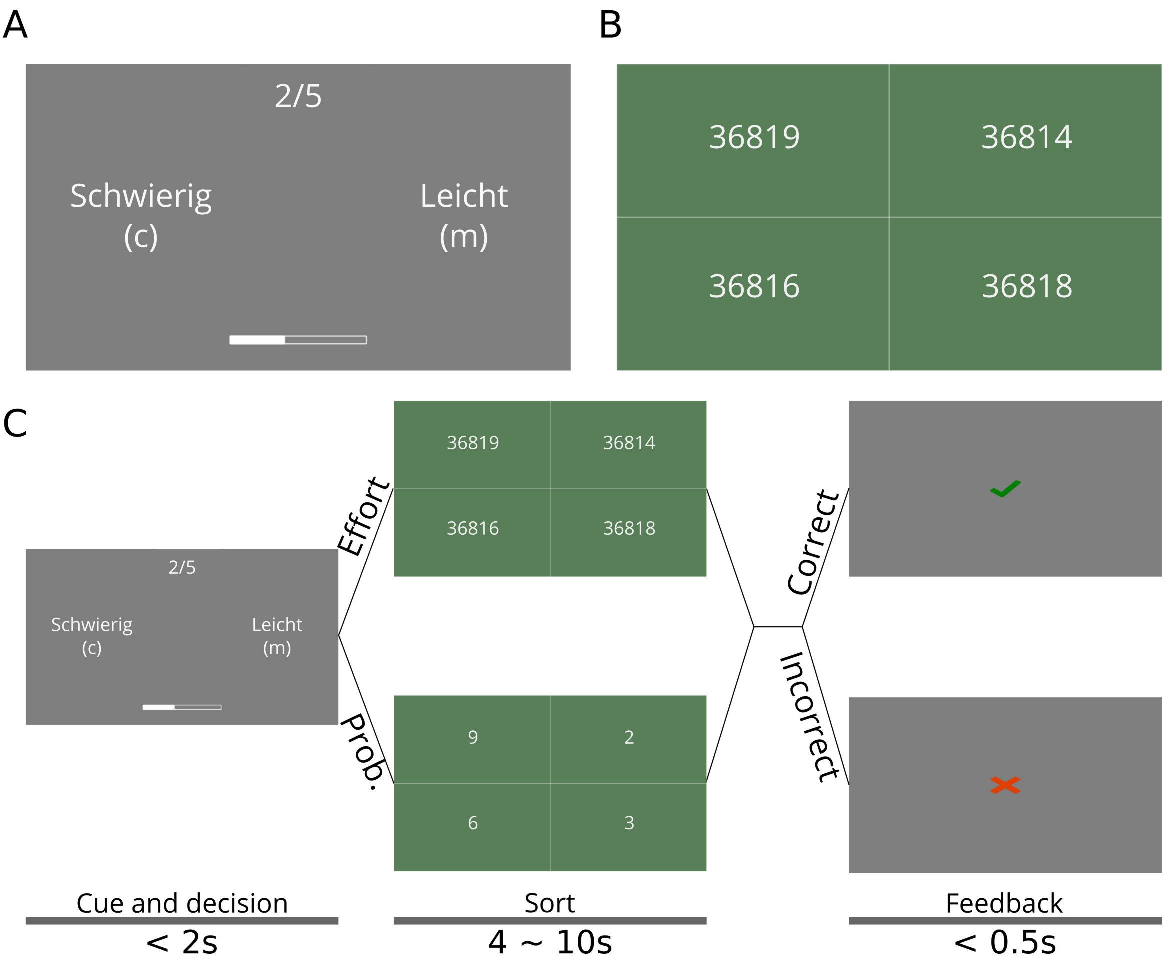

Figure 3.1: Experimental task. **(A)** Cue and decision phase of the sequential task. Participants had to choose between the easy option (Leicht, in the original German), which corresponded to the probabilistic option (see main text), and a hard option (Schwierig), which corresponded to the effortful option, which lead to the task shown in (B). The choice was made with keyboard keys C and M, for the option on the left and right, respectively; the side on which each option appeared is randomly selected at every trial. The trial number is shown on top as 2/5, which means the second trial out of five**(B)** Number-sorting task, where participants had to select the shown numbers in ascending order to correctly complete a single trial of the sequential effort-investment task. To select a single number, participants could click anywhere in the box containing this number. **(C)** Schedule of the different phases of a single trial in the sequential task. The shown screens are those in (A) and (B). The times for each screen are shown at the bottom, along with the name of each phase. Note that the time allotted to sorting the numbers was the same both for the probabilistic (leicht) or effortful (schwierig) option. The main experiment consisted of 25 mini-blocks (sequences of trials) with ten trials each. The text size on panels A and B was increased for visual clarity

trials, divided into six single-digit sorting tasks and six five-digit sorting tasks. Training followed, during which participants' response times for the main experiment were adjusted. Participants first had to go through a block of 40 trials, in which they had to sort the four numbers as quickly as possible within a fixed time-interval of twelve seconds per trial. This was long enough that no participant timed out. After this initial block, the new interval was chosen to be the 95% percentile of the participant's reaction times. After that, three more blocks of 40 trials were possible; after each of them, the participant's performance (i.e., the percentage of times they correctly sorted the numbers before the deadline) was measured. If the performance was below 85%, the deadline was increased. If above 95%, the deadline was decreased. This was repeated for a maximum of four training blocks. If after the training phase the performance was not between 85% and 95%, we excluded the participant from further analysis. The duration of the training phase varied across participants. Once training was done, participants received instructions for the sequential experiment, followed by ten practice mini-blocks, in which they earned no reward (stated in the instructions). Once they finished these, they performed the main experiment with 25 mini-blocks, earning monetary reward for each one completed successfully.

3.3.3 Exclusion criterion

As mentioned in Section 'Procedure', we excluded participants who could not maintain the required accuracy while sorting numbers. The reasoning behind this was twofold. On the one hand, too low performance on the number-sorting task would bias participants towards choosing the effortful option early in the miniblock, to make sure that they had a chance to win. On the other hand, too high performance would leave us unable to adjust the allotted sorting time to ensure that all participants were given time just enough for them to accurately sort the numbers, and no more. A participant that has a 100% accuracy in sorting the numbers may find the task not to be effortful at all.

In total, 5 participants were excluded from analysis due to their success rates being outside of the range 85-95% in the number-sorting task throughout the experiment. The remaining 55 participants were used for the analysis in the Results section.

3.3.4 Single-trial discounting models

The sequential decision-making model proposed in this work is based on classical single-trial discounting models. For completeness, we briefly describe their mathematical form in this section.

It is now well accepted that the best-fitting discounting function for probability discounting is a hyperbola-like one (Ostaszewski et al., 1998), whose mathematical form is given by:

$$\hat{V} = V f_{\mathrm{p}}(p) \tag{3.1}$$

where $\hat{V}$ is the subjective value, p is the probability of obtaining the reward, V is the objective reward value (e.g. the amount of money) and f_{p} is given by:

$$f_{\mathrm{p}}(p) = \frac{1}{\left(1 + \kappa_p \frac{1-p}{p}\right)^s} \tag{3.2}$$

where κ_p and s are the model's free parameters which are to be fit to behavioral data. These two parameters have the effect of creating steeper discounting the higher their values are; κ_p

is regarded as a probability-scaling parameter, while s is regarded as a non-linear sensitivity to probability (Green and Myerson, 2004).

We made use of this model during our study with one caveat: while the inclusion of the parameter s has been previously found to add explanatory power to the model, it makes comparison between groups more difficult (McKerchar and Renda, 2012), as discounting is affected by these two parameters, and it severely complicates parameter fitting due to the high correlation between the parameters (Myerson et al., 2001). For this reason, we chose to fix s to 1 for all participants.

For effort discounting it is less clear which discounting function describes behavioral data best (Białaszek et al., 2017; Klein-Flügge et al., 2015; Kool et al., 2010; Klein-Flügge et al., 2016; Kivetz, 2003). Formal model comparison has been performed between different discount functions, with differing results (Białaszek et al., 2017; Klein-Flügge et al., 2015).

In this work, we exemplify our model using hyperbolic and sigmoid effort discount functions. We chose hyperbolic discounting for its long tradition in probability and delay discounting, which makes it a prime candidate for effort discounting. Sigmoid discounting, on the other hand, has the property of being concave for low effort levels and convex for high effort levels, which Klein-Flügge et al. (2015) argued was an integral part of effort discounting. However, note that our modeling approach presented below can be applied to any other discount function.

The hyperbolic effort discount function is given by:

$$f_\epsilon(\epsilon) = \frac{1}{1 + \kappa_\epsilon \epsilon} \tag{3.3}$$

where ϵ is the effort level and κ_ϵ is the only free parameter, which, as with probability discounting (Equation 3.2), represents effort scaling.

The sigmoid discount function is given by:

$$f_\epsilon(\epsilon) = \left(1 - \left(\frac{1}{1 + e^{-m(\epsilon - \epsilon_0)}} - \frac{1}{1 + e^{m\epsilon_0}}\right)\left(1 + e^{-m\epsilon_0}\right)\right) \tag{3.4}$$

with free parameters m and ϵ_0 that correspond to slope of the function at the center (where the value of the function is 0.5) and the coordinate of the center.

While the interpretation of $\epsilon = 0$ is clear (there is no effort), effort does not have a natural scale like those of delay and probability. Instead, we chose the units of effort such that the effort level of one number-sorting task is $M - 1$, where M is the number of digits of each number to sort. In this scale, the probabilistic option (see Section 'Sequential task') has an effort level of zero and the effortful task has an effort level of four.

3.3.5 Sequential discounting models

In this work we present a novel family of models that bring the single-trial discounting models of the previous section into the realm of sequential decision-making models of goal-directed behavior. To do this, we built on eqs. (3.1) to (3.4) and added a component that implements forward planning over future trials to achieve the goal of filling the point bar during a mini-block.

Action sequences

For our forward-planning model, we first introduce the concept of action sequences π, which we defined as a list of actions to perform in future trials, one for every trial left in the mini-block. Because in the sequential task, the participant must make forced choices between an effortful and a probabilistic option, an action sequence consists of these binary choices, one for each remaining trial until the end of a mini-block. For example, at the very beginning of a mini-block (with ten trials left), an action sequence could consist of only the probabilistic choices at every trial in the future. This would be the policy of a participant who, at the beginning of the mini-block, prefers not to choose the effortful options throughout the mini-block. Another would be an action sequence consisting only of choosing the effortful options. Planning for more nuanced strategies is also possible, i.e. a mix of both options.

The model evaluates every possible action sequence in a way that reflects the overarching goal leading to reward, i.e., filling the point bar. Since at every trial the choice is binary, the total number of possible action sequences at the beginning of trial t is 2^{T-t+1}, including the one to be made at trial t, where T is the total number of trials in a mini-block (ten in our experiment).

It is unlikely that human participants use such a brute-force, binary-tree search algorithm to find the best strategy, as the number of action sequences grows exponentially with the number of trials left; therefore, we created a model in which the only two strategies available are (1) committing to choosing the probabilistic option for the remaining trials in the mini-block and (2) committing to choosing the effortful option for the rest of the mini-block, or until the point bar has been filled. Using only these two action sequences captures the essence of the task, in which a frugal decision-making agent would choose to exert no effort unless it becomes absolutely necessary, and a more reward-sensitive agent (i.e., one that wants to maximize the probability of obtaining reward, disregarding the cost of effort) would prefer exerting effort until the probability of winning the mini-block is high enough to risk the probabilistic option. We discuss the validity and usefulness of this reduction in the number of policies in the Discussion section.

We define these two action sequences with π_p as the action sequence of all-probabilistic choices and π_ϵ as the action sequence of all-effortful choices. With these, we define the set $A = \{\pi_p, \pi_\epsilon\}$.

For every action sequence $\pi \in A$ the model produces an evaluation $z(\pi)$ which determines how beneficial this action sequence is for achieving the goal. Then, the model selects an action (probabilistic or effortful) using these valuations. Concretely, the action a_t at trial t is sampled according to:

$$a_t \sim \sigma_\beta \left(z(\pi_p), z(\pi_\epsilon) \right) \tag{3.5}$$

where σ_β is the softmax function with inverse-temperature parameter β. We fix the value of this parameter to 5 for all models and participants, which produced posterior probabilities (for effort and probability) in the full range of 0 to 1.

The evaluation function z is defined in terms of the single-trial discounting models discussed in Section 'Single-trial discounting models'. In what follows, we discuss $z(\pi_p)$ and $z(\pi_\epsilon)$ separately.

Forward planning with probability

When planning to choose the probabilistic option for every trial into the future, we propose two natural ways of calculating $z(\pi_p)$; one aim of the study was to use model comparison to disambiguate between these two ways. The first way is to stack the discounting function as many times as there are trials left:

$$z(\pi_p) = V f_{\mathsf{p}}(p)^{(T-t+1)} \tag{3.6}$$

$$= V\left(\frac{1}{1+\kappa_p\frac{1-p}{p}}\right)^{T-t+1} \tag{3.7}$$

where $f_{\mathsf{p}}(p)$ is given by Equation 3.2 with $s = 1$. This simply means that the objective reward V obtained at the end of the mini-block is discounted once for each remaining trial. We refer to this variant as "stack". Note that we explicitly do not call this variant 'multiply' because some other discounting functions (not considered in this paper) are not multiplicative.

With the second variant, the model calculates the overall probability of winning the reward by choosing the probabilistic option in every remaining trial in the mini-block, as if it were a single action with an overall probability. The calculation of this overall probability is done with the binomial distribution and the resulting probability is used to apply hyperbolic probability discounting:

$$p_{\mathsf{all}} = \sum_{\hat{x}=X-x}^{\inf} B(\hat{x}, p, T-t+1) \tag{3.8}$$

$$z(\pi_p) = V f_{\mathsf{p}}(p_{\mathsf{all}}) \tag{3.9}$$

$$= V\left(\frac{1}{1+\kappa_p\frac{1-p_{\mathsf{all}}}{p_{\mathsf{all}}}}\right) \tag{3.10}$$

where $B(\hat{x}, p, T-t+1)$ is the probability mass function of the binomial distribution, $\hat{x}$ is the number of successes for the binomial, p is the probability of success and $T-t+1$ is the number of trials left; X is the number of points necessary to win the mini-block and x is the current number of points. $f_{\mathsf{p}}(\cdot)$ is given by Equation 3.2. We refer to this variant as "add".

These two alternative models represent two different ways in which participants could be taking future decisions into consideration. In the "stack" variant, each one of the future trials is seen as an independent probabilistic action, with an associated probability to win and to lose. In contrast, the "add" variant sees all future trials as one single probabilistic action, calculating an overall probability of winning using a binomial distributions. While we do not expect participants to perform such complex calculation, they could calculate an approximation to it and use that to make a decision.

A key difference between these two alternatives is related to the way the discounting curves change as a function of the number of trials left. As can be seen in Figure 3.2A, there is more variability across the curves for the "add" model, all in brown tones, than those for "stack", all in green tones, as the number of points necessary to win (different shades) changes from five (beginning of the miniblock; lightest shades) to one (darkest shades); compared to the "stack" variant, the "add" variant is capable of both discounting less steeply when few points are needed and more steeply when many points are needed. The "stack" model is unable to change the discounting curves as much, for any given value of κ_p.

Forward planning with effort

In analogy to the probabilistic action sequence, we propose two variants of the effortful action sequence evaluation. The first variant is the direct counterpart of the stack variant in

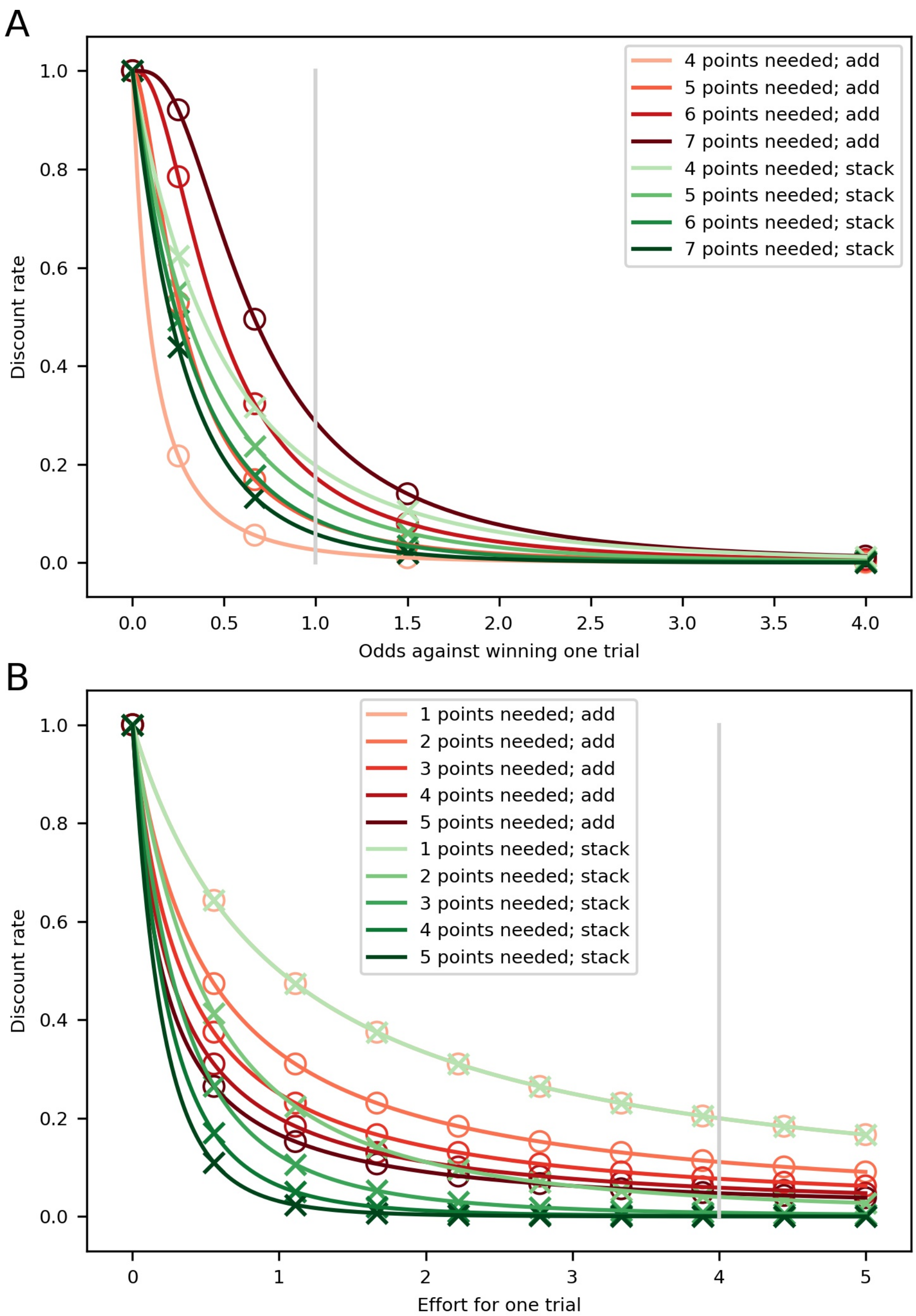

Figure 3.2: Hyperbolic discount curves for the "add" (brown tones) and "stack" (green tones). Markers (circles and crosses) are placed at regular intervals (in probability and effort) only so that it is clear when two lines are overlapping. (A) Probability discounting. Each curve represents a probability discount curve for a number of needed points for the two variants. The parameter κ_p was set to 1 and 5, for the "add" and "stack" variants, respectively. The different values were used so the curves would be as similar as possible. The gray, vertical line represents the 0.5 probability of getting a point used in the experiment. (B) Effort discounting. Each curve represents the effort discount curve for the number of needed points for the "add" (brown tones) and "stack" (green tones) variants. For both variants, the discount parameter was set to $\kappa_e = 1$. The gray, vertical line represents the value used for the effort of a single effortful action.

probability:

$$z(\pi_\epsilon) = V f_\epsilon(\epsilon)^{T-t+1} \tag{3.11}$$

where $f(\epsilon)$ can be hyperbolic effort discounting (Equation 3.3) or sigmoid effort discounting (Equation 3.4). As for the probabilistic action sequence, we refer to this version as 'stack'.

The second variant is the direct counterpart of the add variant in probability, and is defined by adding all the future efforts as if it were a single action and discounting the resulting added effort using the hyperbolic or sigmoid functions:

$$\epsilon_{\text{all}} = (T - t + 1)\epsilon \tag{3.12}$$

$$z(\pi_\epsilon) = V f_\epsilon(\epsilon_{\text{all}}) \tag{3.13}$$

where $f_\epsilon(\cdot)$ can be the hyperbolic effort discounting (Equation 3.3) or sigmoid effort discounting (Equation 3.4). As for the probabilistic action sequence, we refer to this version as 'add'.

As with probability discounting, these two alternative models represent different ways in which the model could consider future effortful actions. In the "stack" variant, as with probability, future effortful actions are seen as independent from present effortful actions and each is discounted separately, which, given that all effortful actions carry the same amount of effort in our task, is represented as a stacking of the discounting function. The "add" variant, on the other hand, posits that effort itself is additive: performing an effortful task takes only half the effort of performing two such tasks. In this variant, the model would think of the effort necessary to perform N effortful actions into the future as having magnitude N times that of a single effortful action and would discount this action sequence based on that added effort.

Additionally, in 3.2B we show how these two models display different behavior. The discount curves change more for the "add" and "stack" variants of the model as fewer future efforts are necessary to win the miniblock. When many points are needed (e.g. at the beginning of the miniblock; darkest shades), the difference between the two variants' discount curves is at its greatest, but as fewer points are needed (close to winning; lightest shades), the curves look more similar between variants until they become the same, as can also be seen from Equation 3.11 and Equation 3.12 by setting $t = T$. It is important to note that the "stack" variant discounts rewards more steeply than the "add" variant when many points are needed, which means that is has a greater range of discounting steepness across the miniblock; this is regardless of the value of the discount parameters. We further discuss these differences in the Discussion section.

Model variants

We defined the different variants of the sequential model depending on the type of forward planning used for effort and probability, each of which could be "stack" or "add". This gave us a total of four variants of the sequential component, naming the effort variant first: add/add, stack/add, add/stack, stack/stack. For example, we refer to the variant in which effort is stacked and probability is added as stack/add.

In addition, two effort discount functions were considered –hyperbolic and sigmoid–, which, combined with the sequential component, yield eight models in total.

3.3.6 Model comparison

In total, we propose a family of eight (2 $\times$ 2 $\times$ 2) models: (sigmoid or hyperbolic) $\times$ (stacking or adding effort) $\times$ (stacking or adding probability). In order to select the one that fits our data best, we implemented the hierarchical model proposed by Stephan et al. (2009), which we only briefly describe here. Note that Stephan et al. (2009) suggest using the so-called exceedance probability to produce a ranking between several models, which takes into account both how many times each model was inferred to be the best for participants, and the uncertainty derived from the inference procedure, making it a more appropriate measure for model comparison than approximations to the model evidence such as the Bayesian information criterion (Schwarz, 1978).

Stephan et al. (2009) defined a hierarchical model in which the models to be compared are first fit to the data of each participant using Bayesian methods. From this fitting, the model evidence can be calculated for every combination of participant and model. This matrix of model evidences is then used as "data" for the hierarchical model. Formally, the model evidence is introduced as $p(d|m)$, where d is the data (participants' choices) and m represents one of the 8 variants we propose, defined as a vector of zeros with a single 1 in the place of the model (for example, the third model is represented by $m = (0, 0, 1, 0, 0, 0, 0, 0)$). This is used to infer, using Bayes theorem, which model best fits the data of all participants together.

The hierarchical model then defines the probability of the model m given an auxiliary variable r:

$$p(m|r) = \prod_{i=1}^{8} r_i^{m_i} \tag{3.14}$$

The variable r_i can be interpreted as the number of participants for which model m_i was the best model (highest model evidence), although this is a simplification. The last component to define is the prior probability of r, which we defined as a flat Dirichlet distribution (as was done by Stephan et al. (2009) in their examples):

$$p(r) = \text{Dirichlet}(\alpha) \tag{3.15}$$

where α is a vector of ones, which reflects that we did not have any hypothesis *a priori* regarding which of the variants of our model fits the data best.

Finally, the full generative model is given by:

$$p(d, m, r) = p(d|m)p(m|r)p(r) \tag{3.16}$$

which we inverted to produce the posterior probability $q(m|d)$ by using the NUTS sampler as implemented in PYMC3 (Salvatier et al., 2016). These posterior distributions can then be used to perform model comparison via the computation of the exceedance probability, which is a way of determine how much more likely is one model to better describe the data than all other models (Stephan et al., 2009).

To calculate the exceedance probability for model i, it suffices to calculate the cumulative distribution of $p(r_i|\text{data})$ over all values for which $p(r_i|\text{data}) > p(r_j|\text{data})$, for all $j \neq i$.

3.3.7 Dividing participants into groups

We divided participants into three groups based on their effort exertion strategy which we determined given their choice data. The first group, called all-effort group, consisted of those participants who chose the effortful option in more than 90% of trials. This implies that these participants used the effortful option even after winning the mini-block.

The remaining participants were divided into two groups: those who applied effort early in the mini-block (early-effort group) and those who applied it late (late-effort group). To divide participants we made use of the frequency of effort calculated at every trial number across mini-blocks. Intuitively, the frequency of effort for participants in the early-effort group decreases as the trial number increases (until the mini-block has been won), while late-effort group increases their frequency with trial number. To quantify this, we calculated the change in frequency of effort between each trial and the next one:

$$m_t = F_{t+1} - F_t, \quad \forall t \in [0, 10) \tag{3.17}$$

where F_t is the overall frequency of effort for trial number t. We found that to classify participants based on when they exerted effort, the best strategy was to count the number of times, for each participant, that the slope was positive for all trials and subtracted the number of times it was negative:

$$\xi_{\text{participant}} = \text{dim}\{t|m_t > 0\} - \text{dim}\{t|m_t \leq 0\} \tag{3.18}$$

where dim() is a function that returns the number of elements in a set. $\xi_{\text{participant}}$ determines whether a participant belongs to the early-effort group ($\xi \leq 0$) or to the late-effort group ($\xi > 0$).

3.3.8 Parameter estimation

Parameter estimation was done using a variational inference scheme implemented the NUTS MCMC sampler implemented in PYMC3 (Salvatier et al., 2016). The outcome of this Bayesian inference scheme is estimations for the mean and standard deviations of the posteriors for each model parameter (see Section 'Sequential discounting models'), providing both a single-point estimate, e.g. the mean of the Gaussian posterior, and estimations for the uncertainty of the inference.

Additionally, the model evidence for all models and participants is calculated as the negative loss produced by PYMC3, which is used for model comparison in Section 'Model comparison'.

Parameter estimation was done using the following generative model:

$$q(\theta|d) = p(d|\theta)p(\theta) \tag{3.19}$$

$$p(\theta) \sim \text{Uniform} \tag{3.20}$$

where $p(\cdot)$ is a probability distribution, θ is the set of parameters to fit to the data and $q(\theta)$ is the posterior distribution over the parameters. Uniform refers to uninformative priors, i.e. prior distributions in which no special prior information is encoded. $p(d|\theta)$ is the likelihood function provided by our decision-making model.

3.4 Results

We first show that there were inter-participant differences in the strategies used to reach the goal, which were reflected in the circumstances under which participants chose the effortful option instead of the probabilistic one. Furthermore, we divided the participants according to three behavioral categories, based on their strategies. This is followed by formal Bayesian model comparison to identify the best among eight different models, which differ in terms of how forward planning computes the subjective value of reward, and which out of two discount functions is used. Having selected the best model for our data, we show that this model correctly captured the overall preference for effort shown by participants. Finally, we show that the overall preference for effort can be understood in terms of the inferred discounting parameters (more specifically, their ratio), providing an intuitive description of apparent effort preference in participants.

3.4.1 Behavioral analysis

Preference for effort

As a first step to determine whether our task elicited differences in the adaptation of effortful choices between participants, we calculated the overall frequency of effort for each participant in the sequential task, i.e. in what percentage of trials the participant chose the effortful option. The results are summarized in Figure 3.3A; to determine whether participants had fully understood the instructions regarding reward contingencies (i.e. that gaining points after filling the point bar would bring no further reward), the trials were separated into before and after having won the mini-block (i.e. filled the points bar), displayed as blue and green bars, respectively. It can be seen that, on average, participants chose the effortful option much less frequently after having won the mini-block, which is congruent with the rules of the task (i.e. that getting more points after having filled the bar is of no use).

In total, we identified three different groups of participants, differing on when they chose to exert effort (see Figure 3.3B and Section 'Dividing participants into groups' in Methods for more details).

We found that 14 (25%) of all participants continued to choose to do effort even after they had won the mini-block. We refer to these participants as the all-effort group for the rest of this work. In the remaining participants we identified two further distinct categories of behavior when looking at those trials before the mini-block had been won, i.e. trials for which the number of obtained points is smaller than five. The first category comprises six (11%) participants that showed a lower frequency of effortful choices at the beginning of the mini-block, averaged across all mini-blocks, and only later increased their frequency. We refer to these participants as the "late-effort" group. The second category, which included 35 (64%) participants, pertains to participants with the opposite behavior; they started every mini-block with a high frequency of effort and only later in the mini-block, when they had accumulated many points (not necessarily having won the mini-block), started choosing the probabilistic option. We refer to these participants as the "early-effort" group.

We considered that all-effort participants may have misinterpreted the instructions of the task. To discard this possibility, we asked all participants in a post-task questionnaire if they understood that gaining points after filling the bar led to no further reward, to which all participants but one responded that they had understood this; the one participant who responded that she did not understand was part of the all-effort group. Importantly, the task was designed such that all participants could easily win all mini-blocks; we found that across all participants, only four mini-blocks were lost (in all cases by a single point) and no participant lost more than one. We will discuss potential reasons for the choice behavior of the all-effort group in the Discussion.

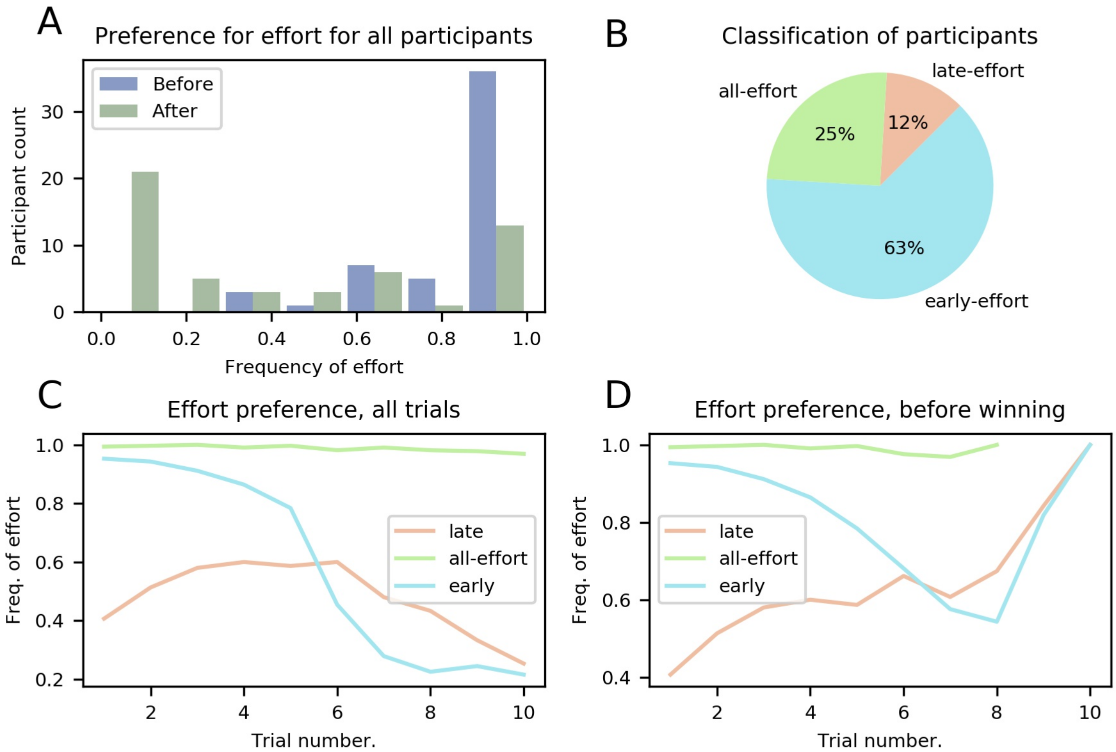

Figure 3.3: **(A)** Histogram of participants' overall frequency of choosing effort averaged across all trials, separated into before (blue) winning the mini-block and after (green). **(B)** Classification of participants into the three groups: all-, early-, and late-effort; see main text. **(C)** Frequency of effort as a function of trial number for the three groups of participants, averaged over participants in each group. **(D)** Same as (C), but only decisions made before the mini-block had been won are included. The different ranges of the lines (e.g. all-effort only reaches trial 8) is because participants who chose effort more often won the mini-block earlier.

The model-based analysis results we present in the following sections can account for the all-effort group simply by inferring very low effort-discounting parameters so that the effortful action no longer comes with disutility and thus can be selected freely. However, the choice data of the all-effort group is rather uninformative about the way individuals resolve the dilemma of when to invest effort to reach a goal that is a few trials away, as one might expect given that they always chose to exert effort. Therefore, the all-effort group will be excluded from the following analyses except when explicitly stated.

The dynamics of the frequency with which participants chose the effortful option can be seen in Figure 3.3C-D for the three categories of participants (late-, early- and all-effort). For Figure 3.3C, we averaged, for every trial number, all the choices made by all the participants in each group. We show in Figure 3.3D the same data but using only the trials before the mini-block had been won. It can be seen clearly that towards the end of the miniblock participants tended to choose to do effort more frequently, because in those mini-blocks when early participants made it to such high trial numbers without having won the mini-block, they urgently needed to accumulate points and thus effort was required to ensure filling the point bar.

3.4.2 Model-based analysis

In this section, we discuss several hypotheses on how exactly human participants select choices in the sequential task. To do this, we use a series of model-based analyses, using Bayesian model comparison to select the best models.

It is important to note that the following analyses are not affected by the distribution of participants across the three groups (early-, late- and all-effort). This is because the models were fitted for each participant separately and the model comparisons are made with all participants.

For all analyses that follow, only trials before the mini-block were used, as only these trials represent goal-seeking behavior.

Forward-planning strategies

We first determined which strategy participants used for forward planning, i.e., how they took into consideration all the possible actions that can be taken in the future and their potential outcomes to decide whether they would exert effort or not at any given trial. Effectively, the question we address here is how the discounting models used to describe single-trial behavior are used by participants in tasks that require forward-planning, goal-reaching behavior.

We considered, for each discounting type (effort or probability), two different ways in which participants computed the subjective value of a reward that can only be obtained after several trials. For future efforts, participants may have used either the strategy to apply the effort discount function as many times as necessary to win the mini-block (we call this “stack”), or adding all necessary efforts to win the mini-block and using the discount function on this sum (we call this “add”). For probability, the strategy can be stacking the discount function (“stack”), or calculating the probability of winning by choosing the probabilistic option all remaining trials (“add”). In total, this resulted in four (two variants for effort $\times$ two variants for probability; for details see Section ’Sequential discounting models’). We refer to the model variants as (effort strategy)/(probability strategy), with the four variants being: add/add, add/stack, stack/add, stack/stack. For example, add/stack refers to the strategy where effort is added

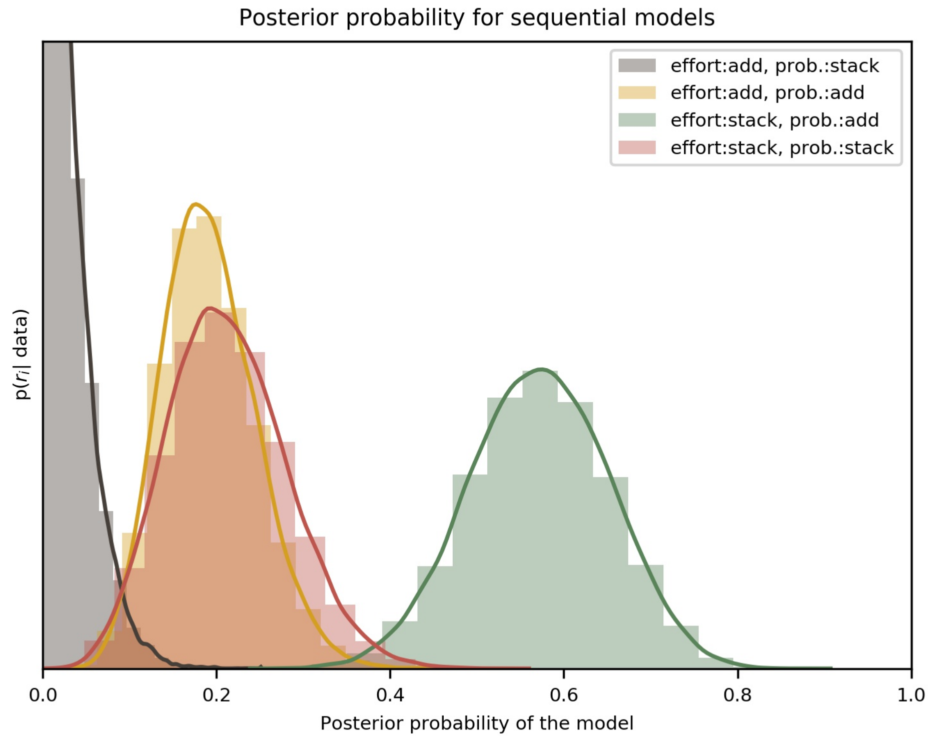

Figure 3.4: The label effort: add/probability: stack, for example, refers to the forward-planning strategy where effort is added and probability stacked. Each of the four distributions, indicated by a solid curve and a histogram, represents the estimated posterior probability of a specific model. It can be clearly seen that the best model for all participants was the 'effort: stack/probability: add' variant. The colored lines are an interpolation with a Gaussian kernel. The two effort discount functions (hyperbolic and sigmoid) have been marginalized to compare only the forward-planning components. The y-axis is the probability density of r_i given the data ($p(r_i|\text{data})$ in Equation 3.14); the x-axis spans all the possible values of r. The peak of the red (add/stack) curve is not shown because the vertical range was cut short for visual clarity.

and probability stacked. To determine which forward-planning strategy was used by participants, we performed formal model comparison between the four forward-planning strategies, following (Stephan et al., 2009).

The results of the model comparison between forward-planning strategies, done by marginalizing over discount functions, can be seen in Figure 3.4. The posterior distributions over the different variants clearly favor the 'effort: stack/probability: add' variant, with an exceedance probability of ˜ 0.99, which means that this forward-planning strategy is orders of magnitude more likely than the others, given the participants' choices.

From our results we can see that the data strongly favors a forward-planning strategy in which future efforts are considered independently of each other (efforts are stacked), discounting the monetary reward at the end of the miniblock once for every future trial in which effort is planned. In contrast, future probabilities are not taken independently; instead, participants seem to calculate the overall probability of winning a miniblock without having to exert any effort and using that calculation for discounting the reward. We further discuss these results in the Discussion section.

Discount functions

Having selected the forward-planning strategy with the highest posterior probability given the data (i.e. effort: stack, probability: add), we set out to determine which effort discount function (sigmoid or hyperbolic) best fit our participants' data. To do this, we performed model comparison between the two discount functions. Our results clearly indicate that hyperbolic effort discounting fits the data better than sigmoid discounting, with an exceedance probability ~1.

These analyses were performed with the data of early- and late-effort participants only, excluding the all-effort group. For completeness, we performed the same analysis including all participants and found that the results do not change. This is due to the fact that, for all models, the effort discounting parameter κ_ϵ (from Equation 3.3) for all-effort participants was estimated to be very low, which caused the model evidence of all models to be the same for that participant. This greatly simplifies model-based data analysis, as it obviates the need for arbitrary exclusion criteria.

Modeling effort preferences

Having selected the best-fitting model for the participants' data (hyperbolic effort discounting, with stacking effort and adding probability, to which we now refer to as HSA), we show in this section that this model indeed captured participants' behavior in a measure not directly used for model comparison: the overall frequency of effort for each participant.

To this end, we compared the HSA model to the experimental data by calculating the overall frequency of effort for each participant across all mini-blocks and doing the same for the models. We performed the analysis only for the early- and late-effort groups. We summarize the results of the comparison in Figure 3.5A, where we show the observed (experimental) and modeled frequencies of effort for each participant separately. We separated the participants into the late- and early-effort groups; the division is shown as a vertical line, to the left of which are the late-effort and to the right, the early-effort participants.

As can be seen in Figure 3.5B, the HSA model estimated the probability of choosing effort very well, being within 5% (in frequency of effortful choices) of the experimental data for most participants. Only for three participants we found an error greater than 15%, which is a level of uncertainty expected from binary data.

It is clear from Figure 3.5A that the fit is better for higher frequencies of effort than for lower. This is because participants with a high frequency of effort have less variability in their choices, which makes them easier to predict by a model. The extreme case of this was participants with an overall frequency of effort (in the early- and late-effort groups) ~ 1, who had almost zero variability in their choices.

Note that for the late-effort group in Figure 3.5A, one participant can be seen with a high frequency of effort. For this participant, effort frequency started very high early in the mini-block and increased as the mini-blocks progressed, meeting our definition of the late-effort group.

Effort allocation

In this section, we show that the overall frequency of effort observed in participants can be explained in terms of the discounting parameters fitted from the HSA model. More specifi-

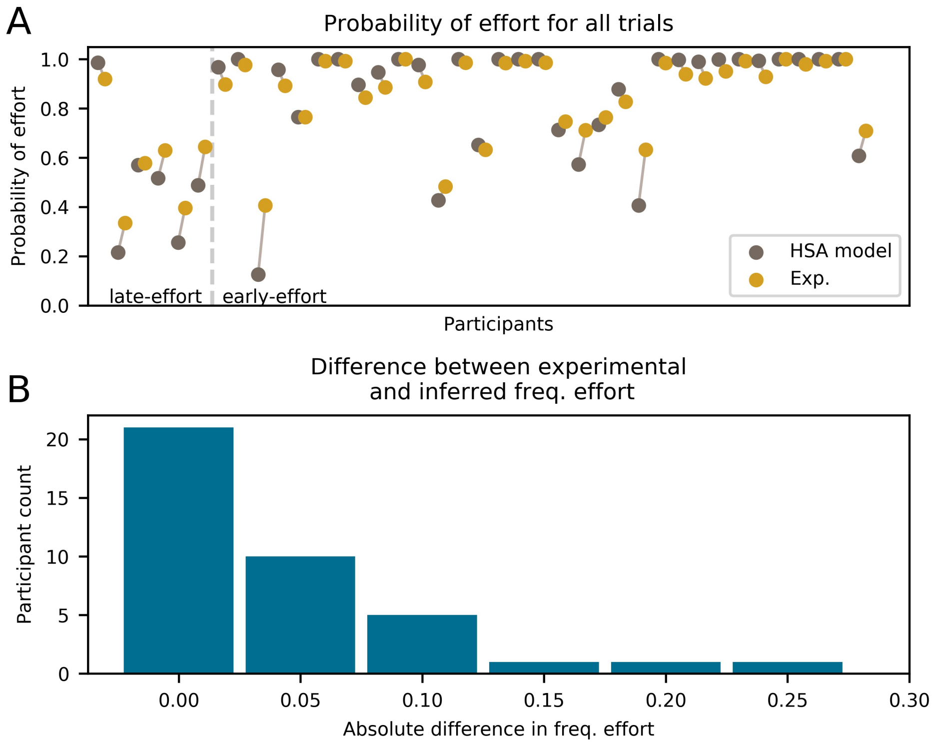

Figure 3.5: Frequency of effort for each participant (excluding the all-effort group) and the HSA model (hyperbolic discounting applied to the 'effort:stack/probability: add' model variant). Only trials before winning the mini-block are included. (A) For each participant, two colored dots are shown, which represent the experimental data (green) and the model prediction (brown). Each dot represents the total frequency of effort for the whole experiment. The two dots for each participant are horizontally offset and connected by a line for visual clarity. Participants are divided by the vertical dashed line into late-effort and early-effort. (B) Histogram of absolute error between the model and the experimental frequency of error shown in (A).

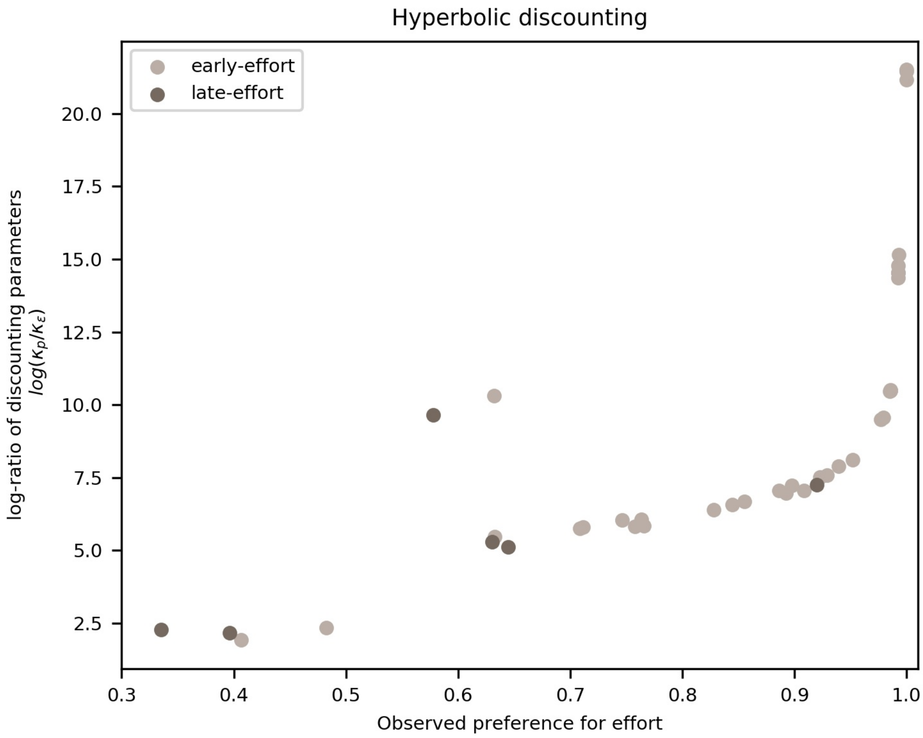

Figure 3.6: Each dot represents a participant, divided into late-effort (dark dots) and early-effort (light dots). We plot the frequency with which a participant chose the effortful action (until reaching the goal of a mini-block) on the x-axis and the log-ratio of the HSA model (hyperbolic discounting applied to the 'effort:stack/probability: add' model variant) parameters for probability to effort discounting, i.e. κ_p to κ_ϵ, on the y-axis (log-scale for clarity).

cally, we show that participants with a higher frequency of effort are those who discounted probability more steeply than effort.

To do this, we calculated, for each participant, the ratio of the posterior means of the HSA model's probability discounting parameter κ_p from hyperbolic probability discounting, to κ_ϵ from effort discounting. Figure 3.6 shows these ratios plotted against the individual overall frequencies of effortful choices. It can be seen that there is a monotonically-increasing relation between the ratio of discount parameters and the overall preference for effort, save for two outliers (one of which has a large absolute difference in Figure 3.5, belonging to the early-effort group).

This monotonically-increasing relation can be interpreted in terms of the comparison between the two options in the task: a participant with a high ratio discounted probability more steeply than effort, which translates into a lower valuation of any probabilistic offer, compared to an effortful one. At values of the frequency of effort ˜1, the log-ratio increases rapidly (faster than exponentially) due to the nature of the model, as the probability of effort grows more slowly than exponentially as κ_ϵ decreases linearly.

3.5 Discussion

We designed a sequential decision-making task in which participants could choose, in each trial, to exert mental effort in order to improve their chances of obtaining reward at the end of a mini-block (i.e., sequence) of ten trials. In this task, participants had the option to exert effort immediately to ensure future reward or choose a probabilistic option and wait until a later trial to re-evaluate if effort needed to be exerted. With this task, we aimed at determining when participants choose to exert effort and which forward-planning strategy they employed to make such a decision. To this end, we proposed a forward-planning model for goal-directed, sequential decision-making behavior that incorporates different strategies for the consideration of future exertion of effort.

Our results show inter-participant variation in when they chose to exert effort, with most participants choosing to start a mini-block with effort and only later choosing not to exert effort. Additionally, the results of our model comparison between four different forward-planning strategies show that most participants considered future efforts by stacking the effort discount function, i.e., by applying the function as many times as they planned to exert effort in future trials. For probability discounting, we found that the best-fitting model calculates the overall probability of reaching the goal (winning a mini-block) when always choosing the probabilistic option. We also found that hyperbolic effort discounting fits the data of our experiment better than sigmoid effort discounting. Finally, we showed that the overall frequency of effort for a participant can be explained by the ratio of the inferred probability discounting to the effort discounting parameters.

3.5.1 Forward-planning strategies

In Section 'Forward-planning strategies' we showed that the forward-planning strategy which best fits the data is one in which effort is "stacked" and probability "added", which we call HSA (for hyperbolic discounting applied to the 'effort:stack/probability: add' model variant). In this model, an overall probability of reaching the goal of the miniblock (i.e. accumulate enough points to fill the bar) is calculated for the all-probability action sequence, and this overall probability is used to discount the monetary reward at the end of the miniblock. In contrast, future efforts are taken into account one at a time, discounting the reward once for every future effort necessary to win the miniblock.

We speculate that this model reflects an important difference in which probabilities and effort are processed by participants. While the probabilities of success of a number of future actions can be collapsed into a single overall probability, this is not done for effort. Rather, effort seems to play a different role in forward planning, whereby a participant asks herself how she will value a reward after each single instance of (future) effort required to obtain it. Such piecemeal considerations could be prompted by the structure of the task itself, where at each future trial, the participant can choose not to continue exerting effort. There could be a difference in the way future effort discounts reward if, instead of having five independent instances of effort exertion, participants could simply choose to exert five times the effort, once. In real life, this would be the difference between having to decide whether to work for five hours in one go, or having to make five sequential decisions to work for one hour, where the decision about each work hour is followed by a prospective, internal evaluation how one will feel, in relation to an overall goal, after having completed one further hour of work.

3.5.2 Future modeling perspectives

Based on previous research, we believe effort and probability are the main driving forces behind behavior in our task. However, this does not preclude the possibility of other effects being in place. Some of these effects could be included in the value of the discount function parameters, like a preference for cognitively-demanding tasks (Westbrook et al., 2013), which we directly infer from the data and therefore implicitly model. Others can be thought of as competing goals in the form of intrinsic motivation, like wanting to please the experimenter, as discussed by Pessiglione et al. (2018), whose effects are constant throughout the experiment, and could be added to the model as part of the reward to be obtained (e.g. with the all-effort action sequence, but not the all-probability one).

A third type of effects comprises dynamic effects, whose influence on decisions changes from trial to trial. In our task, one such effect may be an avoidance of negative feedback, which would differ from extrinsic motivation by reward (obtained when winning a miniblock). We do not believe that such an effect may explain participants' choice because feedback for every trial continues even after the point bar has been filled. This would imply that its effects would need to disappear, or at least be greatly lowered, once enough points have been secured in the miniblock, although winning the miniblock should not affect the desire to avoid negative feedback. However, the effects of feedback could be added as another component of the model, either in the reward space, i.e. that the obtainable reward from the all-effort action sequence is modeled as 20cents + (T - t)(Feedback), where "feedback" is the predicted positive or negative feedback for the action sequence, or as a discounting force, i.e. Subjective Reward = f(negative feedback, monetary reward), where f(., .) is a discount function which decreases with negative feedback.

It would be the subject of future research to determine which of these effects significantly affects behavior to build a more complete account of behavior in such sequential tasks.

3.5.3 Preference for effort

We found that most participants had a strong preference for effort. A quarter of participants (the all-effort group) went as far as choosing to exert effort even when it brought no extra monetary reward. In particular, participants in the all-effort group did not seem to be following the instructions of the task. A similar phenomenon, i.e., continuing to exert effort when it no longer is necessary, has been observed in physical effort experiments (Schmidt et al., 2008; Bouc et al., 2016).

There may be two possible reasons for this phenomenon: First, the level of cognitive effort in our number-sorting task could be too low to trigger a cost/benefit analysis in participants in the all-effort group. In our task, the effortful option came implicitly tied to an increase in the probability of earning monetary reward, which added to the overall benefit of exerting some effort. Moreover, other reasons may be that for some participants, the number-sorting task was interesting on its own (Inzlicht et al., 2018), participants did not want to wait for the next trial while doing nothing, and wanted to make sure they did not lose practice, all of which were reported by our participants in a post-task questionnaire. A related possibility was suggested by Pessiglione et al. (2018), namely that participants might want to "make an impression on the experimenter" by always choosing to exert effort.

Second, we speculate that highly motivated individuals might "flatten" their effort discount curves (e.g., by making κ_ϵ smaller) to more easily attain highly-valued rewards in a scenario like a psychological experiment, which they might misunderstand as a competitive scenario.

As volunteer participants can be assumed to be highly motivated, especially when monetary reward is contingent on performance (Hertwig and Ortmann, 2001), this would mean that their effort discounting parameters are lower, causing the observed high frequency of effort.

Testing these two possible explanations could prove fruitful in future research. Testing the low-effort level possibility would require a task that parametrically varies the effort level to establish higher levels of cognitive effort, as is done typically with physical effort (Prévost et al., 2010). Based on these variations, the proposed model-based approach can be used to infer meta-control by establishing differences in individual effort and probability discounting parameters between different levels of effort requirements.

3.5.4 Action sequences

As part of the present model's definition, we limited the action sequences considered by the model to the all-effort (π_ϵ) and the all-probability (π_p) action sequences (see Section 'Sequential discounting models'). Here, we discuss the reasoning behind this choice and its interesting ramifications.

We posit that as a means to prune the decision tree, participants developed a strategy in which they evaluate the current state of the task and determine it to be "good" or "bad", which in turn allowed them to simplify the decision tree to the two action sequences π_ϵ and π_p. A good state is one in which the participant is close to winning. A bad state is one in which losing seems likely. A good state is then one in which the participant can afford to choose the probabilistic option without it becoming too likely to lose the mini-block, while a bad one is one in which effort needs to be exerted to continue to have a chance at winning. It depends on the participant where exactly this change from good to bad state lies.

In a bad state, effort is, by definition of the bad state, necessary not only in the current trial, but also for all the remaining ones, as otherwise the probabilistic option would still be viable and the state would be good. Therefore considering a mixed action sequence (i.e. one in which both effort and probability can be planned for future trials) is unnecessary in bad states.

In contrast, in a good state, the probabilistic option is still viable. This definition does not preclude future necessity of effort, as things could go wrong and all probabilistic options be lost, which eventually would lead to a bad state. However, as states are evaluated at every trial during the experiment, it is unnecessary to consider this possibility when evaluating the action sequences during a good state; instead, the participant can simply wait until the state has actually become bad in the future and then switch to the all-effort strategy. This implies that good states only require the evaluation of π_p.

How is this state evaluation carried out? Since the only viable option in a good state is π_p and the only viable option in a bad state is π_ϵ, one can turn this around and define a good state as one in which $z(\pi_p) > z(\pi_\epsilon)$, where $z(\cdot)$ is the valuation function (Equation 3.7), and a bad state as one in which the opposite is true. Therefore, the decision-making agent can decide between effort and probability by comparing the valuations of π_p and π_ϵ, as done in the proposed model. This evaluation could be affected by the meta-control we discussed in Section 'Preference for effort'; for example, a highly-motivated individual would classify states as "bad" more often than one with low motivation. Whether motivation and, more generally, meta-control could change which action sequences are evaluated at all should be the target of future research.

3.5.5 Effort and goal reaching

It has been suggested that individuals generally tend to avoid cognitive effort (Kool et al., 2010; Westbrook et al., 2013). However, in the tasks used in the experiments by Kool et al. (2010) and Westbrook et al. (2013), there was no set goal that could be reached more readily via the exertion of cognitive effort. In the study by Kool et al. (2010), participants could not earn additional money if they chose the more effortful task more often. In the experiments in (Westbrook et al., 2013), the association between the actual investment of effort in an increasingly difficult n-back task, the choice behavior in the titration procedure used to determine the subjective value of redoing the different n-back levels, and the actual payment based on four randomly selected choices in the titration procedure may simply have been too unconstrained. In the present task, it was clear in every trial and mini-block that choosing the effortful option would be beneficial for obtaining the reward.

This caveat to the assumption of a general tendency of individuals to avoid the exertion of cognitive effort is also backed by the observation that stable individual differences in personality traits related to the tendency to willingly exert cognitive effort have been found to be associated with effort discounting: Kool and Botvinick (2013) found that individuals with higher scores in Self-Control showed less avoidance of cognitive demand, and Westbrook et al. (2013) observed that participants with higher scores in Need for Cognition showed less effort discounting. While Self-Control is characterized by the investment of mental effort to control one's impulses that interfere with long-term goals (Tangney et al., 2004), Need for Cognition refers to the tendency to engage in and enjoy effortful mental activities (Cacioppo et al., 1996), which can be summarized as cognitive motivation. It remains to be determined whether our participants' habitual cognitive motivation may have played a modulatory role in their decisions to choose the effortful condition more frequently because of their intrinsic motivation to invest cognitive effort. Taken together, our results partly corroborate the seminal findings by Kool and Botvinick (2013) and Westbrook et al. (2013) in pointing to individual differences in the willingness to invest cognitive effort and extend them by showing that the assumption of a general tendency for the avoidance of the exertion of cognitive effort only holds if there is no goal that can be achieved more readily by the exertion of effort.

In conclusion, we have presented a novel combination of a sequential decision making task and a computational model based on discounting effects to describe how participants plan forward to exert effort to reach a goal. We believe that this computational-experimental approach will be highly useful for future studies in the analysis of how participants meta-control the cost/benefit ratio during goal reaching.

4 General discussion

Summary

In this work, I presented two studies on sequential decision making under risk and effort with a strong focus on computational models and their use in model-based data analysis.

In Chapter 2, a computational model of behavior in a sequential task based on the active inference framework (Friston et al., 2015) was described. Fitting this model to each participant's choices enabled the analysis of their behavior with a resolution of single-trial decisions, inferring how certain the participant was that the choice they made was the correct one. In addition to capturing summary statistics with all participants pooled together, as is commonly done in data-based analysis to obtain more reliable statistics, it was shown that the model-based approach is able to do trial-by-trial analysis by using all trials to infer what the certainty of the participant was during each trial.

In Chapter 3, I showed how traditional discounting models, which were made for single-trial decisions, were extended by including a component of forward-planning to model behavior in a sequential task. The purpose of this model was to show that our understanding of probability and effort discounting needs not be revised for tasks outside of the scope of the original experiments (which were single-trial), but instead needs only be extended to the sequential domain. As before, the models were fitted to each participants' choices. I showed that participants' behavior, which adapts to the ever-changing nature of the context in which their choices must be made, can be explained by a set of static discount parameters. Additionally, a comparison was made different possible strategies for taking into consideration future exertion of effort and it was shown that participants are more likely to apply the discount function for every instance of planned effort. It was also showed that the inferred parameter values for each participant correlate with their propensity to wait until later to start investing mental effort into a task, when the evolving environment made it clear that effort had become necessary.

4.1 Applications

A theme common to the two presented studies is that of model-based data analysis. As the name suggests, it is the idea of analyzing the behavioral data through the lens of a computational model, which provides a more in-depth look into the decisions being analyzed. In the case of the models presented here, the most important output is the probability distribution from which each decision is sampled in the model. These distributions encode the certainty with which an agent, mimicking a participant, makes a decision. Importantly, this information is not directly accessible in the choice data, and to approximate it with traditional analysis methods, many different trials must be pooled together to obtain reliable statistics, e.g. by pooling all participants' decisions together.

Such summary statistics are not without merit. Indeed, a crucial step in establishing the validity of the model-based analysis was to show that it can reproduce these summary statistics. In both studies discussed above, I showed that the model-based analysis captured elements of the participants' behavior that were not directly used to fit the models, which is formally known as posterior predictive checks (PPC; Gelman et al., 1996). This can be seen in Figure 2.7 and 3.5, where the posterior probabilities given by the models for each trial are averaged and compared to the decisions made by individual participants and by all participants, respectively.

Beyond reproducing traditional analyses, model-based data analysis provides additional insights. In this work, two different types of insights are presented and discussed: (1) In Chapter 2, the fitted model was used to essentially extend the available data by calculating the posterior probability of an action in situations (contexts) which were not observed during the experimental session. This extrapolation answers the question: what would the participant do in this new context? (2) In Chapter 3, the fitted parameters from the models were used directly to describe a trend in the behavior of all participants which cannot be seen from the data alone. While the latter form has seen many uses, e.g. with discounting parameters (Shamosh and Gray, 2008; Jullien et al., 1999; Myerson et al., 2003), to my knowledge, the former has seen limited use in psychology or neuroscience due to the difficulty of building and verifying an adequate behavioral model.

The most straightforward application of model-based data extrapolation is that of fitting a model with (abundant) training data in a behavioral experiment and using the fitted model to predict and study the behavior of participants during an imaging (e.g. fMRI) session, in which the number of trials might not be enough to reliably fit the model. In this situation, the extrapolated data is not different from the data used for fitting, as the participant is typically performing the same task (with variations due to, for example, stochasticity). However, a more involved application of model-based extrapolation is that of predicting and understanding the behavior of people in situations in which participants cannot be placed during an experiment due, for example, to potential danger to the participant, or to time constraints in the performance of the task (e.g. when comparing behavior against a control condition in pharmacological interventions that can only be applied for a short duration).

In addition to the two tools discussed above, model-based analysis also provides an extra tool for analysis: the posterior distribution over actions for each trial. These distributions were used in all model-based analyses in Chapter 2 and Chapter 3. In this sense, the fitted models can be seen as a tool that maps the binary choices of a participant to the certainty with which the participant made these choices, in a trial-by-trial basis. This could enable trial-by-trial analyses of, e.g., the neural underpinnings of the preference for risk which go beyond a binary regressor (as done in Schwartenbeck et al., 2015) essentially cleaning up the noise inherent to forced binary choices from the behavioral measures (e.g. Kolling et al., 2014). Further development of these methods is of paramount importance in the future, as directly obtaining these certainties from participants is difficult in most cases, downright impossible in some.

4.2 Limitations and future work

4.2.1 Model validation

The power and reliability of model-based analyses naturally depend on the validity of the model. Because of this, it is of great importance to find ways to ensure that the model accurately simulates human behavior or, put another way, find and apply tools to determine how well the model fits the available data and predicts unavailable data.

While for some types of models (e.g. linear models) known measures of model fit exist which can be interpreted on their own, the generic case of a non-linear model with arbitrary form is more complicated.

If a standard model already exists, and a new, possibly better model is presented, model comparison can be used to determine whether the new model better fits the data than the incumbent one. In all other cases, like those presented in Chapter 2 and Chapter 3, one form of model validation left is that of posterior predictive checks (PPCs). As discussed above, PPCs were used to validate the models used in the presented studies.

PPCs take the form of summary statistics that are highly dependent on the data being used to fit the models. In the case of participant choice data, participant-level statistics like overall performance on the task (e.g. number of accumulated points), average reaction times (e.g. in the drift diffusion model (Ratcliff and McKoon, 2008)) or propensity for one choice over the others (as discussed in Chapter 2 and Chapter 3) make for simple PPCs which are easily interpretable.

A more general PPC, which could potentially be applied to most behavioral experiments, is that of comparing the posteriors over actions to the certainty reported by participants. In future works, an essential step in the introduction of a new behavioral model could be to directly validate a subset of these posteriors. To do this during a behavioral task, the experimenter need only ask the participant how certain she was about the previous decision. While this can be disruptive and, in some cases, impractical, for the purpose of PPC, this could be done only for a small but significant subset of trials. The decisions in these trials can be excluded from all analysis, and the reported certainties used exclusively as a PPC.

These reported certainty levels can be compared against the posterior probabilities over actions produced by the fitted model to ascertain whether there is a close fit. Having done this, we can be sure not only that the model accurately captures human behavior, but also that the posterior probabilities themselves can be taken at face value.

4.2.2 Implicit vs explicit modeling of risk

In the two models presented, there is a significant difference in how risk is modeled. While the models of Chapter 3 (henceforth, the explicit models) incorporate a specific component of probability discounting (i.e. risk discounting), the active inference-based models (henceforth, the implicit models) do not, instead relying on an elaborate forward-planning machinery which culminates in the evaluation of sequences of actions using a goal function.

In the explicit models, risk aversion is inferred for each participant and the interpretation of the inferred values can be made as has been made in single-trial experiments (e.g. Green and Myerson, 2004): if the parameter κ_p (see Equation 3.2) is bigger than one, the participant is risk-averse. The opposite for $0 < \kappa_p < 1$, and if $\kappa = 1$, the participant follows the expected value.

On the other hand, implicit models do not lend themselves to such interpretation. In these models, how much the agent should accept risk is determined based on how many points are needed, as well as on the goal shape used by the participant. In a sense, risk aversion is calculated optimally for each trial, given the (subjective) goal shape for the participant. As such, one cannot say that a participant is more or less risk averse, but rather that she is more or less accepting of lower-point outcomes (of the miniblock): for example, an agent with very flat goals (see Figure 2.2) is not as concerned with making it past the threshold, and thus could choose the safe option more often.

The difference between the two types of models goes beyond mathematical definitions. The implicit models posit that risk aversion does not exist in isolation, but is rather a consequence of both the goals and the predictions during forward planning. In this perspective, a person does not simply prefer risk (e.g. riskier bets with higher payoffs), but instead finds sub-maximal outcomes (e.g. when the reward at the end is not as high as it could have been) unacceptable, making risk a necessity. The explicit models, on the other hand, posit that risk aversion is an independent mechanism from that of reward evaluation (although affected by it (Estle et al., 2006)), manifesting as a direct preference for safer bets.

Whether risk aversion and the value function are separate entities is not a new question. Dyer and Sarin (1982) argued that what might appear as risk aversion can be better explained in terms of a nonlinear evaluation of possible rewards and what they call relative risk aversion. In contrast, prospect theory Kahneman and Tversky (1979); Tversky and Kahneman (1992) explicitly separates the effect of risk aversion and the value function.

The high adaptability of implicit models could provide a mechanistic explanation for the finding that inferred discount parameters are affected by, among others, reward magnitude, gains vs. loss, and context. However, establishing such models has proven to be difficult, as they have, by necessity, many components, each of which would need to be verified independently. The importance of this question cannot be overstated and, as such, I believe it should be the focus of future studies.

Bibliography

Anderson, J. R. (2000). *Learning and Memory: An Integrated Approach, 2nd Ed*. Learning and Memory: An Integrated Approach, 2nd Ed. John Wiley & Sons Inc, Hoboken, NJ, US.

Apps, M. A. J., Grima, L. L., Manohar, S., and Husain, M. (2015). The role of cognitive effort in subjective reward devaluation and risky decision-making. *Scientific Reports*, 5:16880.

Ashby, F. G., Maddox, W. T., and Lee, W. W. (1994). On the Dangers of Averaging Across Subjects When Using Multidimensional Scaling or the Similarity-Choice Model. *Psychological Science*, 5(3):144–151.

Averbeck, B. B. (2015). Theory of Choice in Bandit, Information Sampling and Foraging Tasks. *PLoS Comput Biol*, 11(3):e1004164.

Basile, A. G. and Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions, thinking dispositions, and behavioral outcomes. *Frontiers in Psychology*, 6.

Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1):23–36.

Białaszek, W., Marcowski, P., and Ostaszewski, P. (2017). Physical and cognitive effort discounting across different reward magnitudes: Tests of discounting models. *PLOS ONE*, 12(7):e0182353.

Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: Drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, 8:102.

Bornstein, A. M., Khaw, M. W., Shohamy, D., and Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, 8:ncomms15958.

Botvinick, M. and Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, 16(10):485–488.

Botvinick, M. M., Huffstetler, S., and McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, 9(1):16–27.

Bouc, R. L., Rigoux, L., Schmidt, L., Degos, B., Welter, M.-L., Vidailhet, M., Daunizeau, J., and Pessiglione, M. (2016). Computational Dissection of Dopamine Motor and Motivational Functions in Humans. *Journal of Neuroscience*, 36(25):6623–6633.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., and Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2):197–253.

Camerer, C. F. and Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19(1):7–42.

Caraco, T., Blanckenhorn, W. U., Gregory, G. M., Newman, J. A., Recer, G. M., and Zwicker, S. M. (1990). Risk-sensitivity: Ambient temperature affects foraging choice. *Animal Behaviour*, 39(2):338–345.

Caraco, T., Martindale, S., and Whittam, T. S. (1980). An empirical demonstration of risk-sensitive foraging preferences. *Animal Behaviour*, 28(3):820–830.

Cartar, R. and Dill, L. (1990). Why are bumble bees risk-sensitive foragers? *Behavioral Ecology and Sociobiology*, 26(2):121–127.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007). Consistency and Heterogeneity of Individual Behavior under Uncertainty. *American Economic Review*, 97(5):1921–1938.

Clarke, A. C. (1968). *2001: A Space Odyssey*. New American Library, New york.

Cohen, A. L., Sanborn, A. N., and Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, 15(4):692–712.

Cuevas Rivera, D., Ott, F., Markovic, D., Strobel, A., and Kiebel, S. J. (2018). Context-Dependent Risk Aversion: A Model-Based Approach. *Frontiers in Psychology*, 9.

Cuevas Rivera, D., Strobel, A., Goschke, T., and Kiebel, S. J. (2020). Modeling dynamic allocation of effort in a sequential task using discounting models. *Frontiers in Neuroscience*, 14.

Daunizeau, J., Friston, K. J., and Kiebel, S. J. (2009). Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D: Nonlinear Phenomena*, 238(21):2089–2118.

Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, 5(4):613–624.

Dolan, R. J. and Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, 80(2):312–325.

Dyer, J. S. and Sarin, R. K. (1982). Relative Risk Aversion. *Management Science*, 28(8):875–886.

Economides, M., Guitart-Masip, M., Kurth-Nelson, Z., and Dolan, R. J. (2015). Arbitration between controlled and impulsive choices. *NeuroImage*, 109:206–216.

Estes, W. K. and Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, 12(3):403–408.

Estle, S. J., Green, L., Myerson, J., and Holt, D. D. (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Memory & Cognition*, 34(4):914–928.

Everitt, B. J. and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 8(11):1481–1489.

Fantino, E., Gaitan, S., Kennelly, A., and Stolarz-Fantino, S. (2007). How reinforcer type affects choice in economic games. *Behavioural Processes*, 75(2):107–114.

FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., and Friston, K. (2015). Active inference, evidence accumulation and the urn task. *Neural computation*, 27(2):306–328.

Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator:L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4):453–476.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 0(0):1–28.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1655):20130481.

Gallagher, K. M. and Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, 43(1):101–116.

Gelman, A., Meng, X.-l., and Stern, H. (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. *Statistica Sinica*, pages 733–807.

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: Advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, 23 Suppl 1:41–57.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F. Å., and Hansen, L. K. (1999). On Clustering fMRI Time Series. *NeuroImage*, 9(3):298–310.

Green, L. and Myerson, J. (2004). A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological bulletin*, 130(5):769–792.

Green, L., Myerson, J., Lichtman, D., Rosen, S., and Fry, A. (1996). Temporal discounting in choice between delayed rewards: The role of age and income. *Psychology and Aging*, 11(1):79–84.

Green, L., Myerson, J., and Ostaszewski, P. (1999a). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 25(2):418–427.

Green, L., Myerson, J., and Ostaszewski, P. (1999b). Discounting of delayed rewards across the life span: Age differences in individual discounting functions. *Behavioural Processes*, 46(1):89–96.

Hartmann, M. N., Hager, O. M., Tobler, P. N., and Kaiser, S. (2013). Parabolic discounting of monetary rewards by physical effort. *Behavioural Processes*, 100:192–196.

Hertwig, R. and Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *The Behavioral and Brain Sciences*, 24(3):383–403; discussion 403–451.

Hey, J. D. and Orme, C. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica*, 62(6):1291–1326.

Houston, A. I., Fawcett, T. W., Mallpress, D. E. W., and McNamara, J. M. (2014). Clarifying the relationship between prospect theory and risk-sensitive foraging theory. *Evolution and Human Behavior*, 35(6):502–507.

Hull, C. (1943). *Principles of Behavior*. Appleton-Century, Oxford, England.

Inzlicht, M. and Berkman, E. (2015). Six Questions for the Resource Model of Control (and Some Answers). *Social and personality psychology compass*, 9(10):511–524.

Inzlicht, M., Shenhav, A., and Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in cognitive sciences*, 22(4):337–349.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). *SciPy: Open Source Scientific Tools for Python*.

Jullien, B., Salanié, B., and Salanié, F. (1999). Should More Risk-Averse Agents Exert More Effort? *The Geneva Papers on Risk and Insurance Theory*, 24(1):19–28.

Kacelnik, A. and Bateson, M. (1996). Risky Theories—The Effects of Variance on Foraging Decisions. *Integrative and Comparative Biology*, 36(4):402–434.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Kaplan, R. and Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, 112(4):323–343.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kellen, D., Pachur, T., and Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition*, 157:126–138.

Kivetz, R. (2003). The Effects of Effort and Intrinsic Motivation on Risky Choice. *Marketing Science*, 22(4):477–502.

Klein-Flügge, M. C., Kennerley, S. W., Friston, K., and Bestmann, S. (2016). Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *Journal of Neuroscience*, 36(39):10002–10015.

Klein-Flügge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D., and Bestmann, S. (2015). Behavioral Modeling of Human Choices Reveals Dissociable Effects of Physical Effort and Temporal Delay on Reward Devaluation. *PLOS Computational Biology*, 11(3):e1004116.

Kolling, N., Behrens, T. E. J., Mars, R. B., and Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, 336(6077):95–98.

Kolling, N., Wittmann, M., and Rushworth, M. F. S. (2014). Multiple Neural Mechanisms of Decision Making and Their Competition under Changing Risk Pressure. *Neuron*, 81(5):1190–1202.

Kool, W. and Botvinick, M. (2013). The intrinsic cost of cognitive control. *Behavioral and Brain Sciences*, 36(6):697–698.

Kool, W., McGuire, J. T., Rosen, Z. B., and Botvinick, M. M. (2010). Decision Making and the Avoidance of Cognitive Demand. *Journal of experimental psychology. General*, 139(4):665–682.

Kurzban, R., Duckworth, A., Kable, J. W., and Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, 36(6):661–679.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Locey, M. L., Jones, B. A., and Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and decision making*, 6(6):552–564.

Mather, M., Mazar, N., Gorlick, M. A., Lighthall, N. R., Burgeno, J., Schoeke, A., and Ariely, D. (2012). Risk preferences and aging: The "certainty effect" in older adults' decision making. *Psychology and Aging*, 27(4):801–816.

Mather, M., Shafir, E., and Johnson, M. K. (2003). Remembering chosen and assigned options. *Memory & Cognition*, 31(3):422–433.

McKerchar, T. L. and Renda, C. R. (2012). Delay and probability discounting in humans: An overview. *The Psychological Record*, 62(4):817–834.

McNamara, J. M. and Houston, A. I. (1992). Risk-sensitive foraging: A review of the theory. *Bulletin of Mathematical Biology*, 54(2):355–378.

Mobbs, D., Hassabis, D., Yu, R., Chu, C., Rushworth, M., Boorman, E., and Dalgleish, T. (2013). Foraging under Competition: The Neural Basis of Input-Matching in Humans. *The Journal of Neuroscience*, 33(23):9866–9872.

Moran, R., Pinotsis, D. A., and Friston, K. (2013). Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, 7.

Myerson, J., Green, L., Scott Hanson, J., Holt, D. D., and Estle, S. J. (2003). Discounting delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology*, 24(5):619–635.

Myerson, J., Green, L., and Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, 76(2):235–243.

Nasiry, J. and Popescu, I. (2012). Advance Selling When Consumers Regret. *Management Science*, 58(6):1160–1177.

O'Brien, E. L. and Hess, T. M. (2020). Differential focus on probability and losses between young and older adults in risky decision-making. *Aging, Neuropsychology, and Cognition*, 27(4):532–552.

Ostaszewski, P., Green, L., and Myerson, J. (1998). Effects of inflation on the subjective value of delayed and probabilistic rewards. *Psychonomic Bulletin & Review*, 5(2):324–333.

Pedersen, M. L., Frank, M. J., and Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, 24(4):1234–1251.

Pessiglione, M., Vinckier, F., Bouret, S., Daunizeau, J., and Le Bouc, R. (2018). Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases. *Brain*, 141(3):629–650.

Phillips, P. E. M., Walton, M. E., and Jhou, T. C. (2007). Calculating utility: Preclinical evidence for cost–benefit analysis by mesolimbic dopamine. *Psychopharmacology*, 191(3):483–495.

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., and Dreher, J.-C. (2010). Separate Valuation Subsystems for Delay and Effort Decision Costs. *Journal of Neuroscience*, 30(42):14080–14090.

Radulescu, E., Nagai, Y., and Critchley, H. (2015). Mental Effort: Brain and Autonomic Correlates in Health and Disease. In Gendolla, G. H., Tops, M., and Koole, S. L., editors, *Handbook of Biobehavioral Approaches to Self-Regulation*, pages 237–253. Springer, New York, NY.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34(6):1446–1465.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55.

Schmidt, L., d'Arc, B. F., Lafargue, G., Galanaud, D., Czernecki, V., Grabli, D., Schüpbach, M., Hartmann, A., Lévy, R., Dubois, B., and Pessiglione, M. (2008). Disconnecting force from money: Effects of basal ganglia damage on incentive motivation. *Brain*, 131(5):1303–1310.

Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., and Pessiglione, M. (2012). Neural Mechanisms Underlying Motivation of Mental Versus Physical Effort. *PLOS Biology*, 10(2):e1001266.

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., and Friston, K. (2015). The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. *Cerebral Cortex*, 25(10):3434–3445.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Schweighofer, N., Shishida, K., Han, C. E., Okamoto, Y., Tanaka, S. C., Yamawaki, S., and Doya, K. (2006). Humans Can Adopt Optimal Discounting Strategy under Real-Time Constraints. *PLOS Computational Biology*, 2(11):e152.

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N. R., Rajeshwar, S., de Brebisson, A., Sotelo, J. M. R., Suhubdy, D., Michalski, V., Nguyen, A., Pineau, J., and Bengio, Y. (2017). A Deep Reinforcement Learning Chatbot. *arXiv:1709.02349 [cs, stat]*.

Shamosh, N. A. and Gray, J. R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence*, 36(4):289–305.

Shead, N. W. and Hodgins, D. C. (2009). PROBABILITY DISCOUNTING OF GAINS AND LOSSES: IMPLICATIONS FOR RISK ATTITUDES AND IMPULSIVITY. *Journal of the Experimental Analysis of Behavior*, 92(1):1–16.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., and Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, 40(1):99–124.

Siegler, R. (1987). The Perils of Averaging Data Over Strategies - an Example from Childrens Addition. *Journal of Experimental Psychology-General*, 116(3):250–264.

Skvortsova, V., Palminteri, S., and Pessiglione, M. (2014). Learning To Minimize Efforts versus Maximizing Rewards: Computational Principles and Neural Correlates. *Journal of Neuroscience*, 34(47):15621–15630.

Steel, P. and König, C. J. (2006). Integrating theories of motivation. *The Academy of Management Review*, 31(4):889–913.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, 46(4):1004–1017.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA.

Talmi, D. and Pine, A. (2012). How Costs Influence Decision Values for Mixed Outcomes. *Frontiers in Neuroscience*, 6.

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, 72(2):271–324.

Toussaint, M. and Storkey, A. (2006). Probabilistic inference for solving discrete and continuous state Markov Decision Processes. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 945–952, Pittsburgh, Pennsylvania, USA. Association for Computing Machinery.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ, US.

Walasek, L. and Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General*, 144(1):7–11.

Walton, M., Kennerley, S., Bannerman, D., Phillips, P., and Rushworth, M. (2006). Weighing up the Benefits of Work: Behavioral and Neural Analyses of Effort-Related Decision Making. *Neural networks : the official journal of the International Neural Network Society*, 19(8):1302–1314.

Wang, L., Zheng, J., and Meng, L. (2017). Effort provides its own reward: Endeavors reinforce subjective expectation and evaluation of task performance. *Experimental Brain Research*, 235(4):1107–1118.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Westbrook, A., Kester, D., and Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLOS ONE*, 8(7):e68210.

Winterhalder, B. and Smith, E. A. (2000). Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evolutionary Anthropology: Issues, News, and Reviews*, 9(2):51–72.

Xu, S., Pan, Y., Qu, Z., Fang, Z., Yang, Z., Yang, F., Wang, F., and Rao, H. (2018). Differential effects of real versus hypothetical monetary reward magnitude on risk-taking behavior and brain activity. *Scientific Reports*, 8(1):3712.

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; Hilfe dritter wurde nur in wissenschaftlich vertretbarem und prüfungsrechtlich zulässigem Ausmaß in Anspruch genommen, die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Es sind keine unzulässigen geldwerten Leistungen, weder unmittelbar noch mittelbar, im Zusammenhang mit dem Inhalt der vorliegenden Dissertation an Dritte erfolgt. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die vorgelegte Dissertation wurde unter wissenschaftlicher Betreuung von Prof. Stefan Kiebel am an der Professur für Neuroimaging der Technischen Universität Dresden erstellt.

Es haben keine früheren erfolglosen Promotionsverfahren stattgefunden.

Die Promotionsordnung des Bereichs Mathematik und Naturwissenschaften der Technischen Universität Dresden, in der Fassung vom 23.02.2011, letzte Änderung 23.05.2018, wird anerkannt.

Unterschrift | Datum