UNIVERSITÄTSKLINIKUM
HAMBURG-EPPENDORF

Zentrum für Experimentelle Medizin
Institut für Systemische Neurowissenschaften
Prof. Dr. med. Christian Büchel

Dopaminergic modulation of the explore/exploit trade-off in human decision making

Dissertation

zur Erlangung des Doktorgrades Dr. rer. biol. hum.
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von
Karima Chakroun
aus Bremerhaven

Hamburg 2019

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 27.06.2019

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende:	Prof. Dr. Jan Peters
Prüfungsausschuss, zweite/r Gutachter/in:	Prof. Dr. Steffen Moritz
Dritte/r Gutachter/in:	Prof. Dr. Gerhard Jocham

Datum der Disputation: 27.06.2019

Contents

1 Introduction

1.1 The explore/exploit trade-off

1.1.1 The concept of explore/exploit

Our lives are made up of countless decisions. These range from relatively trivial ones, such as which ice cream to buy or which shirt to wear, to important and potentially life-altering decisions about which career to pursue and which partner to choose. A central aspect of many decision problems is the regulation of when to exploit, i.e. to choose a familiar option with a well-known reward, and when to explore, i.e. to try an alternative option with an unknown or uncertain but potentially higher reward. This decision dilemma is commonly known as the "explore/exploit trade-off" and is encountered in many different life situations (for reviews see Addicott, Pearson, Sweitzer, Barack, & Platt, 2017; Cohen, McClure, & Yu, 2007). To get a better understanding of the concept and complexity of explore/exploit problems, consider the following example: Suppose you are planning your next summer vacation and need to decide where to travel. On the one hand, the safe choice would be to spend the holidays at your favorite and well-known holiday resort in Italy, which you have visited and enjoyed for the past five years. On the other hand, there are also some other promising holiday destinations like France or Portugal, which you have never visited before. How should you decide? Arriving at an optimal decision for this problem is not easy, since different aspects have to be considered. Since you only go on summer vacation once a year and spend much hard earned money on it, you want it to be a most rewarding experience. While you already know that a holiday in Italy is always very rewarding, you are still uncertain about how enjoyable a trip to one of the other countries would be. Yet, exploring one of the unknown alternatives has both its advantages and disadvantages. On the upside, exploration helps you to gather information about the alternatives and reduce decision uncertainty in the following years. Also, this information could be useful for maximizing rewards in the long term, as you might find out that there are holiday destinations that even surpass Italy and are hence more worthwhile to exploit. Moreover, exploring different alternatives from time to time becomes even more important if the rewarding qualities of different choice options change over time, such as certain countries becoming less attractive for vacations while others start to bloom. On the downside, exploring one of these unknown alternatives is more risky in its outcome and only comes at a cost: the time and effort spent on planning and travelling, the money paid for flight tickets and hotels, as well as the "opportunity costs", i.e. the rewards forgone by not exploiting your favorite holiday resort in Italy. Also, excessive exploration may lead to unnecessary losses and deplete resources like money and time without returning much reward. Hence, striking a good balance between exploration and exploitation is essential in order to maximize rewards and minimize costs in the long term (Addicott et al., 2017) and can be regarded as a "fundamental need for adaptive behaviour in a complex and changing world" (Cohen et al., 2007, p. 934).

The explore/exploit dilemma is a very common problem faced by all kinds of decision makers. Humans, for instance, encounter this decision trade-off at various scopes and timescales, ranging from everyday choices like exploiting your favorite meal vs. exploring a new dish on the menu, to once-in-a-life decisions like staying in your old profession vs. exploring alternative career paths. Similarly, also animals encounter the explore/exploit dilemma when foraging for limited resources like food, shelter, or mates (Addicott et al., 2017; Cook, Franks, & Robinson, 2013; Mehlhorn et al., 2015). For example, a honey bee that feeds on a flower's nectar has to decide how long to exploit that flower (or patch of flowers) and when to move on to the next (Katz & Naug, 2015). Even foraging microorganisms have been shown to face and solve the explore/exploit dilemma, for example when growing out in different directions before concentrating growth at a particular area of high nutritional payoff (Reid et al., 2016; Watkinson et al., 2005; see also Cohen et al., 2007). Furthermore, explore/exploit trade-offs are not only encountered by single subjects, but are also relevant for larger instances, e.g. for organizational learning (Gupta, Smith, & Shalley, 2006; Lavie, Stettner, & Tushman, 2010; March, 1991), business management (Molina-Castillo, Jimenez-Jimenez, & Munuera-Aleman, 2011; Uotila, Maula, Keil, & Zahra, 2009), or scientific and cultural innovation systems (Hills, Todd, Lazer, Redish, & Couzin, 2015). Finally, explore/exploit trade-offs are not limited to living agents, but often need to be solved by computer algorithms applied in machine learning and artificial intelligence, e.g. for swarm robotics (Alers et al., 2011; Baldassano & Leonard, 2009), web content optimization (Agarwal, Chen, & Elango, 2009), and user recommendation systems (Lacerda, Santos, Veloso, & Ziviani, 2015; Mahajan, Rastogi, Tiwari, & Mitra, 2012). Taken together, all these diverse examples demonstrate that the explore/exploit dilemma is a widespread and fundamental part of various real-world decision problems.

Despite the high occurrence and relevance of the explore/exploit trade-off across several fields, research is still at the beginning to understand how humans and non-human agents solve this dilemma. In fact, many important questions revolving around this topic have remained unanswered so far, some of which will be briefly considered here before addressing more specific research findings on the explore/exploit trade-off in the following sections. A first important and open question is how agents should optimally solve the explore/exploit dilemma. Until now, there is no known optimal solution for this decision problem and it is unclear if such an optimal solution even exists (see Cohen et al., 2007). Empirical research has focused instead on describing how living organisms actually behave when facing explore/exploit problems and on studying the mechanisms underlying this behavior (e.g. Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Lee, Zhang, Munro, & Steyvers, 2011; Steyvers, Lee, & Wagenmakers, 2009). A second question still under debate is whether exploration and exploitation are two qualitatively different and competing processes, or if they are better conceptualized as extreme ends of a continuum (see Addicott et al., 2017; Gupta et al., 2006; Mehlhorn et al., 2015). According to the continuum idea, behaviors at the extreme ends may be disadvantageous, considering that too much exploitation fosters inflexibility and habit formation, whereas too much exploration may lead to

volatile, inefficient decision making and prevent the formation of expertise (Addicott et al., 2017; Beeler, Cools, Luciana, Ostlund, & Petzinger, 2014). Optimal behavior may occur somewhere in between both extremes, where exploration and exploitation are balanced. Third, as the explore/exploit trade-off extents to numerous disciplines and can be assessed by various paradigms (see 1.1.2), it is a very heterogeneous construct that still lacks a clear definition and unified theoretical framework so far (Hills et al., 2015; Mehlhorn et al., 2015). This heterogeneity of definitions and paradigms makes it difficult to integrate findings from different studies and research fields, and also poses a problem for the overall reproducibility of research on the explore/exploit trade-off (Addicott et al., 2017; Helversen, Mata, Samanez-Larkin, & Wilke, 2018). For example, the current literature offers at least three different approaches for defining exploration and exploitation (see Mehlhorn et al., 2015). These approaches are either based on (a) observable behavioral patterns, i.e. staying (exploit) vs. switching (explore), (b) expected rewards, i.e. choosing the option with the highest expected reward (exploit) vs. choosing an alternative option (explore), or (c) obtained outcomes, i.e. obtaining explicit rewards like money or food (exploit) vs. obtaining information (explore). Adding to the complexity, these three approaches are by no means mutually exclusive, and most explore/exploit concepts involve more than one of them (Mehlhorn et al., 2015). Only recently, scientists have started to synthesize research from different fields for building a general and multidisciplinary framework of the explore/exploit trade-off (Berger-Tal, Nathan, Meron, & Saltz, 2014; Mehlhorn et al., 2015). Lastly, research is only beginning to reveal the neural mechanisms underlying explore/exploit decisions, i.e. the brain regions and neurotransmitter systems involved therein. Although some progress has been made in this direction over the past decade (see 1.1.4 and 1.1.5), many aspects remain poorly understood and further studies are needed to gain more knowledge on the neural substrates of explore/exploit behavior in health and disease (see 6.6).

Following this broad introduction, the next sections dig deeper into the existing literature on the explore/exploit trade-off with a focus on behavioral and neuroscientific research in humans and animals. First, the main paradigms for assessing explore/exploit behavior are presented (1.1.2), along with an overview of the most common cognitive modeling approaches taken in this field (1.1.3). Then, recent research on the brain regions (1.1.4) and neurochemical systems (1.1.5) involved in explore/exploit behavior is reviewed, before specifically focusing on the dopaminergic system (1.2) and its potential role in the explore/exploit trade-off (1.3).

1.1.2 Paradigms to study explore/exploit behavior

Several behavioral paradigms have been developed to study explore/exploit decisions in animals and humans. In the following, four of the most commonly used paradigms are described in more detail, as these are reencountered later in this introduction and in the discussion.

The most widely used behavioral paradigm to study explore/exploit decisions in the laboratory is the n-armed (or multi-armed) bandit task (Gittins, 1979; Gittins & Jones, 1974; originally described by

Robbins, 1952). The basic idea of this task is that of a casino's slot-machine known as the one-armed bandit, with the difference that the n-armed bandit task offers several (n) arms to be pulled, each paying out at a different and initially unknown reward rate. The number of arms varies between studies, but often bandit tasks with two to four arms are used. In each trial, the subject chooses one bandit, which then reveals its payoff before a new trial starts. The overall goal of the game is to maximize the overall payout. Depending on the bandits' payoff structures, different types of bandit tasks can be distinguished. First, bandit tasks can be distinguished into a binary and non-binary version. In the binary version, each bandit pays out binary rewards, i.e. a fixed reward or no reward, with a certain reward probability. In contrast, bandits in the non-binary version pay out continuous rewards like points or cents that vary randomly (e.g. normally distributed) around a certain mean reward value. Second, and more importantly, bandit tasks can be distinguished into stationary and non-stationary (restless) bandit problems. In the stationary version, reward rates of each bandit are fixed throughout the task, meaning that the reward probability (for binary rewards) or the reward mean value (for continuous rewards) does not change over trials (e.g. Harlé et al., 2015; Steyvers et al., 2009). In this case, subjects need to learn the expected reward of each bandit only once in order to choose the one with the highest expected payoff. In the restless bandit task, however, reward rates (i.e. reward probabilities or mean payoffs) of each bandit vary slowly and randomly over time, challenging subjects to repeatedly choose between exploiting the currently best option and exploring unfamiliar options to keep track of their changing reward rates. Research on the explore/exploit trade-off is typically based on the restless bandit paradigm to study how subjects balance exploration and exploitation in such dynamic environments (although see below for the horizon task). For instance, Daw et al. (2006) used a restless four-armed bandit task with continuous payoffs to study explore/exploit behavior and its neural correlates in human subjects. Since the same task was also used in the current project, a detailed description of its procedure can be found in the methods section (see 2.4).

The restless bandit paradigm offers several advantages for research. First, it is easily applicable to both humans and animals, allowing to study explore/exploit behavior in and across different species (Addicott et al., 2017). For example, bandit tasks have been adapted for monkeys (Costa, Tran, Turchi, & Averbeck, 2014; Pearson, Hayden, Raghavachari, & Platt, 2009), pigeons (Racey, Young, Garlick, Pham, & Blaisdell, 2011), mice (Naudé et al., 2016), and even microorganisms (Reid et al., 2016). Second, in its basic characteristics, the bandit problem is representative of a broad class of real-world decision problems encountered in dynamic environments, some examples being food foraging, partner search, consumer decisions, and organizational learning (see Hills et al., 2015; Mehlhorn et al., 2015). Third, the multi-armed bandit problem is well amenable to formal analysis and has already been extensively studied in the field of reinforcement learning (e.g. Berry & Fristedt, 1985; Brezzi & Lai, 2002; Gittins, 1979; Kaelbling, Littman, & Moore, 1996; Macready & Wolpert, 1998; Sutton & Barto, 2018). As a result, a wide repertoire of elaborate algorithms and cognitive modeling approaches

already exists for this problem, which can be profitably applied to empirical studies in order to examine explore/exploit behavior and its underlying cognitive and neural processes more deeply (see 1.1.3).

A second widely used paradigm to study explore/exploit decisions is the (patch) foraging task, which exists in various forms. Foraging tasks use a more naturalistic scenario, in which animals or humans explore and exploit sources of food, either in a real life setting (e.g. Cook et al., 2013; Evans & Raine, 2014; Hall, Humphries, & Kramer, 2007; Latty & Beekman, 2013) or in a virtual environment (e.g. Addicott, Pearson, Kaiser, Platt, & McClernon, 2015; Lenow, Constantino, Daw, & Phelps, 2017; Mata, Wilke, & Czienskowski, 2013). For example, Constantino and Daw (2015) used a virtual foraging task, in which human subjects are presented with an apple tree and have to decide whether to harvest that tree or to move on to a new and unharvested one. Crucially, exploiting a tree entails only a short harvest delay (3 s), but leads to an exponential depletion in the amount of harvested apples, while exploring a new tree entails a much longer travel time delay (9 s). Other human studies have used similar virtual foraging tasks with comparable scenarios, e.g. with berry bushes (Addicott et al., 2015) or fishing ponds (Hutchinson, Wilke, & Todd, 2008; Mata et al., 2013). All these tasks essentially measure how long subjects exploit a current source of food before abandoning it to search for a better one, a decision also known as "patch leaving" (Hayden, Pearson, & Platt, 2011; Hutchinson et al., 2008; Pyke, 2018; Stephens & Krebs, 1986). The term "patch" therein refers to the fact that natural food sources often occur in unevenly distributed clumps or patches (e.g. patches of flowers, fruits on a tree), such that leaving a depleting patch to explore a new one is often associated with a certain traveling cost. While the foraging paradigm shows similarities to the restless bandit problem, it also differs in some important aspects. In the typical foraging task, different patches or options can only be chosen one after another, such that a previously exploited option cannot be visited again and each explored option has never been visited before. Also, exploitative decisions result in a successive reduction of the reward rate due to depletion, while reward rates in the bandit task are independent of subjects' choices. Hence, both paradigms offer slightly different approaches to measure explore/exploit decisions and it eventually depends on the research question, which of these tasks is more suitable and more representative of the real-world setting under study.

A third paradigm that has often been used to study explore/exploit decisions in humans is the clock task (Moustafa, Cohen, Sherman, & Frank, 2008; see also Frank, Doll, Oas-Terpstra, & Moreno, 2009). In this task, subjects see a clock face with one arm that rotates once around the clock over the course of five seconds. Subjects are instructed to stop the arm at any time before it makes a full turn, whereby the time of response affects the number of points won. After stopping the arm, subjects receive feedback about the number of rewarded points before the new trials starts. The entire task includes four separate blocks à 50 trials, each block with a different latent reward function that determines how reward magnitude and probability change with response time. Similar to the bandit task, the clock task challenges subjects to balance the exploration of different choice options (here: response times) with the exploitation of the option with the highest expected reward. However, due to the fixed reward

structure within each block, the task prompts transitions from exploration to exploitation only once during each block, and transitions from exploitation to exploration only in between blocks (similar to a stationary bandit task). Also, subjects are explicitly instructed for each block to "try to respond at different times along the clock cycle" (i.e. to explore) in order "to learn how to make the most points" (i.e. exploit; Moustafa et al., 2008, p. 12295). Thus, how subjects solve the explore/exploit-trade-off is less flexible and more pre-determined by the task structure and instructions than in the restless bandit task. A second limitation of the clock task relates to the fact that the classification of a choice as exploratory is based on the observation of a large response time difference between successive trials within a task block (see Frank et al., 2009), which is more likely to arise from decision noise than exploratory choices (switches to another option) in the bandit and foraging task (Addicott et al., 2017).

In addition to these three paradigms, there are some other tasks that have been used to study the explore/exploit trade-off (e.g. Blanchard & Gershman, 2018; Glass et al., 2011; Knox, Otto, Stone, & Love, 2012; Navarro, Newell, & Schulze, 2016; Wilson, Geana, White, Ludvig, & Cohen, 2014). Yet, basically all these paradigms share a set of common features, as reviewed by Addicott et al. (2017). First, there are multiple options to choose from (simultaneously or sequentially). Second, these options are associated with initially unknown reward rates and need to be sampled in order to learn their current reward rates and predict their future outcomes. Third, each decision is a trade-off between exploiting a familiar option and exploring a less familiar option to reduce uncertainty, albeit for an opportunity cost of forgoing the option with the highest known immediate reward.

Aside from these classical paradigms to measure explore/exploit behavior, a new variant of the bandit task has recently been developed (Wilson et al., 2014), which allows to distinguish between different types of explorations. This new variant is called the "horizon task", since it implements different time horizons for decision making. The horizon task works analogous to a stationary two-armed bandit-task, offering two choice options with different reward rates that are fixed within each task block. In contrast to a typical bandit task, however, each block in the horizon task starts with four forced-choice trials, during which subjects just observe the choices and their outcomes. These forced-choice trials either provide equal (two vs. two) or unequal (one vs. three) information about the two choice options. After that, the block continues with either one (short horizon) or six (long horizon) free-choice trials, whereby the horizon of each block is visible to the subjects. This manipulation of the time horizon creates two different settings for the decision maker: one in which the focus is only on the immediate decision (short horizon), and one in which early exploration might pay out in later choices (long horizon). Modeling behavior in the different task conditions (short vs. long horizon, equal vs. unequal information) then allows to quantify the extent to which subjects use random exploration arising from decision noise versus directed exploration, which is selectively targeted towards information seeking and uncertainty reduction. Using this task, Wilson et al. (2014) found that both decision noise and information seeking are increased with longer time horizons and concluded that humans use both random and directed exploration to solve the explore/exploit dilemma. Since then, the horizon task

and variants thereof have successfully been applied in further human studies to differentiate between random and directed exploration (e.g. Cogliati Dezza, Yu, Cleeremans, & Alexander, 2017; Krueger, Wilson, & Cohen, 2017; Somerville et al., 2017; Zajkowski, Kossut, & Wilson, 2017). One limitation of the horizon task is, however, that it is rather time inefficient, since only the first free choice in each block, thus only < 15 % of all trials, are used in the behavioral modeling analysis, compared to 100 % of trials in the restless bandit paradigm. Another limitation relates to the fact that the horizon task is formulated as a series of discrete games, in which reward-related information is reset prior to the start of each new game (see Cogliati Dezza et al., 2017). Hence, explore/exploit behavior can only unfold over very limited time frames in this task (1-6 trials), whereas it unfolds in the restless bandit task over the course of a whole experiment (e.g. 300 trials), making the latter more suitable for studying dynamic transitions between both decision strategies over longer time frames.

1.1.3 Cognitive modeling of explore/exploit behavior

Most explore/exploit paradigms – including the bandit task, clock task, and horizon task – rely on cognitive modeling of the observed behavioral data. Cognitive models provide a mathematical description of the processes underlying subjects' behavior, from which quantifiable parameters can be inferred for subsequent analyses. Modeling of behavior offers in general many advantages over using only raw behavioral data. First, modeling can largely promote the understanding of the cognitive and neural mechanisms generating the observed behavior, since it requires a clear and complete formal characterization of all modeled processes. Within such formalizations, all variables and their dependencies need to be fully specified before the model is implemented, leaving no room for vagueness or incoherence in its theoretical assumptions (Fum, Missier, & Stocco, 2007). Such a clear and formal specification furthermore facilitates scientific communication of a theory and aids reproducibility (Anderson, 2014). Second, mathematical models generally allow for precise behavioral predictions that can be directly tested against the observed data to evaluate the quality of a model. Thereby, different candidate models can be quantitatively compared (e.g. based on their predictive accuracies or goodness of fit) to arrive at increasingly accurate formal descriptions of behavioral phenomena, a method called "quantitative model selection" (Burnham & Anderson, 2010; Lewandowsky & Farrell, 2011; Zucchini, 2000). In this regard, cognitive models can be viewed as tools for scientific discovery, since they enable scientists to investigate the implications of different theoretical ideas "beyond the limits of human thinking" (McClelland, 2009, p. 16). Third, the quantification of model parameters by model fitting can reveal (subtle) behavioral differences between experimental conditions or between normal and clinical populations that may not manifest in model-free behavioral variables (see e.g. Addicott et al., 2017; Harlé et al., 2015). Such model parameters can moreover be used as individual difference measures with regard to a certain behavior or underlying process in order to examine how these relate to other psychological constructs, such as personality traits or intelligence (e.g. Steyvers et al., 2009), or to aid psychiatric diagnostics and research (Adams, Huys, & Roiser, 2016; Stephan & Mathys, 2014; Wang & Krystal, 2014). Moreover,

cognitive models can be used for a trial-by-trial analysis of experimental data (Daw, 2011), allowing to estimate otherwise subjective quantities like the expected reward for every single trial and choice option in a bandit task experiment. This point is especially relevant for research on explore/exploit behavior, since the classification of choices as exploitative or exploratory is most commonly based on model-derived estimates of the expected reward value (see Daw et al., 2006; Mehlhorn et al., 2015). That is, a choice is classified as exploitative if the option with the highest expected reward is chosen, whereas all other choices are classified as exploratory. Combined with functional neuroimaging, this trial-by-trial analysis also makes it possible to examine the neural correlates of these model-derived quantities (Gläscher & O'Doherty, 2010; O'Doherty, Hampton, & Kim, 2007). This approach, called "model-based neuroimaging" (see 2.8.1), has already provided a substantial contribution to understanding the neural mechanisms underlying learning and decision making (see reviews by Daw & Doya, 2006; Dreher, 2013; O'Doherty, 2004). Finally, formal models enable scientists to develop and evaluate mathematically optimal solutions to particular behavioral problems. For instance, optimal decision strategies for trading-off exploration and exploitation have already been formally developed for stationary bandit problems with finite or infinite horizons (Averbeck, 2015; Berry & Fristedt, 1985; Gittins & Whittle, 1989; Kaelbling et al., 1996; Lee et al., 2011), whereas the restless bandit problem still remains unsolved (see Cohen et al., 2007).

In the following, an overview of the cognitive models most commonly applied in empirical research on explore/exploit behavior is given, with a main focus on the bandit paradigm. For a more extensive review of the various modeling approaches used in this field, the reader is referred to the computational modeling and machine learning literature (e.g. Berry & Fristedt, 1985; Daw, 2014; Dayan & Sejnowski, 1996; Gittins, Glazebrook, & Weber, 2011; Kaelbling et al., 1996; Sutton & Barto, 2018; Thrun, 1992).

Modeling choice behavior in the multi-armed bandit task is mostly based on models of reinforcement learning (RL; Sutton & Barto, 1998, 2018). Reinforcement learning is a type of machine learning that formally describes how agents learn the expected reward values of different choice options and how they use this knowledge to select actions so as to maximize the overall reward. A key aspect in these models is the process of error-driven learning, which entails that an agent can only learn about the rewards by taking actions and observing their outcomes. Each action thereby results in the computation of a reward prediction error, quantifying the difference between the received and expected reward of that action, which is then used to update the expected reward for the next trial. An important parameter in most RL models is the learning rate, which determines the degree to which expectations are updated by the reward prediction error. A classical RL model for updating is called the "Delta rule" (Sutton & Barto, 1998), for which a formal description can be found in the methods section (see 2.7.2). Aside from the Delta rule, research on explore/exploit behavior has also often applied the "Bayesian learner" model for updating (e.g. Daw et al., 2006; Speekenbrink & Konstantinidis, 2015). This model is based on the same principle of error-driven learning as the Delta

rule, but uses a different parametrization for the learning process. The Bayesian learner model includes no learning rate parameter, but instead a set of parameters to quantify an agent's assumptions about the underlying reward distributions of different choice options and (in a restless bandit paradigm) how these change over time. This parametrization allows to compute trial-by-trial estimates for both the expected mean and variance of each reward distribution, thereby offering a way to model subjects' uncertainty about the learned values in contrast to the simpler Delta rule. A formal description of the Bayesian learner model can also be found in the methods section (see 2.7.2).

Update rules (learning models) like the Bayesian learner or the Delta rule describe the process by which subjects learn the expected rewards of different choice options, but not the processes by which subjects select their actions based on this knowledge and by which they explore. The process of action selection is instead modeled by choice rules (decision models), whereby different choice rules can be used to implement different explore/exploit strategies, which range from simple heuristics to complex mathematical models.

One of the simplest choice rules applied to (two-armed) bandit problems with binary outcomes is the learning-independent Win-Stay Lose-Shift (WSLS) heuristic (Robbins, 1952). According to this heuristic, a subject continues to choose an option if it returned a reward (win-stay), but switches to the other option if it returned no reward (lose-shift). In a stochastic variant of the WSLS heuristic, subjects stay after winning or shift after losing only with a certain probability, which is determined by the model parameter γ (e.g. Steyvers 2009; Harle 2015). However, the WSLS heuristic is by no means a sophisticated decision strategy, as it only uses reward information of the current trial for selecting the next action, disregarding all the information gained by previous trials. Another simple choice rule is the ε-greedy strategy and different variants thereof (Sutton & Barto, 1998). The ε-greedy strategy is often used within the RL framework to model random exploration behavior in an otherwise greedy (exploiting) agent. This strategy assumes that on each trial, the subject explores with a small probability of ε by choosing randomly from all options, and with a probability of $1 - \varepsilon$ exploits the option with the highest expected reward. Thereby, the ε parameter of the model controls the balance between exploration and exploitation, which stays constant over time and typically adopts small values like $\varepsilon = 0.1$ (i.e. 10 % probability to explore; Sutton & Barto, 2018; Vermorel & Mohri, 2005). One variant of this rule is the ε-first strategy (Even-Dar, Mannor, & Mansour, 2002; Vermorel & Mohri, 2005). Here, exploration and exploitation are assumed to occur in two distinct and subsequent stages. For the first εT trials (where T denotes the total number of trials), the subject only explores by choosing randomly between all options, while for the remaining trials, the subject exploits the option with the highest expected reward. A second variant of the ε-greedy strategy is the ε-decreasing rule (Sutton & Barto, 1998; Vermorel & Mohri, 2005), in which the exploration rate decreases over trials. Formally, this strategy starts with an exploration probability of ε_0 in the first trial, which gradually declines to an exploration probability of ε_0 / i in the ith trial. Hence, this modification of the ε-greedy model allows the explore/exploit balance to shift over the course of learning. Together, the ε-greedy strategy and

its variants all belong to a class called “semi-uniform strategies” (see Vermorel & Mohri, 2005), which have in common that they imply a binary distinction between greedy behavior, in which the best known option is always exploited, and random exploration, in which the choice probability is uniformly distributed across all remaining options.

Aside from these simple decision strategies, one of the most widely applied choice rules in the field of RL is the softmax rule (McFadden, 1974; Sutton & Barto, 1998; see 2.7.2 for formula). According to this rule, choice behavior can vary gradually between pure exploitation and pure exploration, which is controlled by the softmax (β) parameter. A β of zero reflects a purely exploratory behavior with equal choice probability for all options, whereas an extremely large β reflects a purely exploitative (greedy) behavior, in which the best known option is always taken. In between those extremes, choices are probabilistically based on the relative expected rewards of all available options, whereby the inverse β reflects the noisiness of the probabilistic decision (see also 2.7.2). That is, options with a larger expected value have a higher probability to be chosen, but also inferior options can still be selected due to decision noise. In this way, the softmax function allows to model choice behavior as a combination of both exploitative (value-driven) and exploratory (noisy) choice tendencies. The softmax rule is typically applied with a β parameter that is constant over trials (e.g. Daw et al., 2006). In dynamic variants of the softmax rule, however, the β parameter can increase or decrease over trials according to different mathematical functions (e.g. Cesa-Bianchi & Fischer, 1998; Speekenbrink & Konstantinidis, 2015; Vermorel & Mohri, 2005), thereby allowing the explore/exploit balance to change over the course of learning.

Both the softmax rule and semi-uniform strategies like ε-greedy describe exploration based on decision randomness and thus only capture undirected (random) exploratory behavior. However, it has been argued in the literature that one important goal of exploration is to gather information and reduce uncertainty for future choices (Averbeck, 2015; Cogliati Dezza et al., 2017; Dayan & Sejnowski, 1996; Payzan-LeNestour & Bossaerts, 2012; Wilson et al., 2014). For this reason, exploratory choices might not be fully random, but (at least partly) directed towards options with more uncertain outcomes, for which exploration will be most informative. To capture this kind of directed exploration, other choice rules have been developed that are based on the use of an exploration (or information) bonus (e.g. Cogliati Dezza et al., 2017; Daw et al., 2006; Dayan & Sejnowski, 1996; Sutton, 1990; Wilson et al., 2014). The general idea behind these choice rules is to introduce an extra value (bonus) into the model that increases with uncertainty and biases choices towards options with more uncertain outcomes. For instance, Daw et al. (2006) used a modification of the softmax rule, called the “softmax with exploration bonus”, to model directed exploration in the restless bandit task. According to this model, choices are probabilistically based on the sum of a bandit’s mean expected reward plus an exploration bonus (see 2.7.2 for a formal description of this model). This exploration bonus, in turn, is computed on each trial from the variance (uncertainty) of an option’s expected reward rate, which is tracked for each bandit by the Bayesian learner model (see above). A key parameter of the modified

model is the exploration bonus parameter φ, which determines how strong this uncertainty is weighted in the softmax rule. Hence, the φ parameter reflects the degree to which the uncertainty of an outcome, relative to the expected value of an outcome, influences choice behavior. A similar approach has been taken by Cogliati Dezza et al. (2017) for modeling directed exploration in the horizon task. Their model, called "knowledge RL model", includes a two-fold learning rule, which not only learns reward values (via the Delta rule), but also information values for all options. The information value of an option is learned as a function of previous observations, i.e. it increases each time the option is chosen. Then, both types of values are combined by subtracting the (weighted) information value from the expected reward value, thereby devaluating options that have been observed more often in the past. This combined reward-information value is then used, in place of the simple expected reward value, in the standard softmax rule for action selection. Although the knowledge RL model (Cogliati Dezza et al., 2017) and the softmax with exploration bonus (Daw et al., 2006) use a different parametrization, both models capture directed exploration behavior which is driven towards actions with more uncertain outcomes. While these and similar approaches have often been used to model directed exploration, empirical evaluation of these models has yielded inconclusive results so far. While Daw et al. (2006) found no evidence in support of an exploration bonus in human decision making, studies on the horizon task concluded that humans use indeed a combined reward-information value for solving explore/exploit problems (e.g. Cogliati Dezza et al., 2017; Wilson et al., 2014). Yet, these studies differed in many aspects, from the behavioral paradigm to the modeling approach taken, and hence their results are difficult to compare. Note also that a more detailed elaboration on these controversial findings is provided in the discussion (see 6.5.2).

Aside from the mentioned choice rules, further and more complex models of explore/exploit behavior have been developed. While a detailed description of all these models exceeds the scope of this introduction, some of their main ideas are briefly presented. For instance, a model called the "Knowledge Gradient" (Frazier, Powell, & Dayanik, 2008; Harlé et al., 2015; Zhang & Yu, 2013) has been applied to stationary bandit problems, which implements the idea that choices are based on a combination of an immediate reward gain and a long-term knowledge gain. This knowledge gain approximates the value of exploration in each trial, i.e. the degree to which collecting information may pay off in the future. Crucially, the trade-off between reward and knowledge gain changes with the distance to the horizon, such that exploitation is increasingly favored over exploration with fewer trials left (i.e. with an approaching horizon). Another sophisticated choice rule that has been applied to the restless bandit task is the "Probability of Maximum Utility" (PMU) model (Speekenbrink & Konstantinidis, 2015). According to this model, an agent takes into account the whole expected reward distribution of each bandit when making decisions, which is tracked by the Bayesian learner model (see above). In each trial, these reward distributions are quantitatively compared between all options in a pairwise manner to determine the probability (density) of each option to pay off the maximal reward, which is then used as that option's choice probability. The PMU model has some similarities

to the softmax rule with exploration bonus, but does not require an extra parameter to model exploratory decisions. Rather, explore/exploit decisions in the PMU model follow naturally from the learned reward distributions, which are used to maximize the expected reward in each trial. Note that the PMU model therefore implies that uncertain options are not chosen to reduce uncertainty (for exploratory purposes), but because they are expected to have the highest probability to yield the largest immediate reward (for exploitative purposes), in contrast to other models of uncertainty-driven exploration. An empirical evaluation in human subjects has shown that the PMU model outperformed the softmax rule with or without an exploration bonus for many subjects, whereas the standard softmax rule still fitted best across all subjects (Speekenbrink & Konstantinidis, 2015). In addition to the PMU model, other choice rules have been developed that take into account the uncertainty associated with each choice option to model exploration (Audibert, Munos, & Szepesvári, 2009; Auer, Cesa-Bianchi, & Fischer, 2002; Lai & Robbins, 1985). For example, "Upper Confidence Bound" (UCB) models use the estimated uncertainty (e.g. the variance) to compute the upper confidence bound for all options before choosing the option with the highest such bound (see Audibert et al., 2009). Finally, some cognitive models of explore/exploit behavior assume that the agent does not strictly follow one particular choice rule, but rather adapts its decision strategy to changes in the environment (Ishii, Yoshida, & Yoshimoto, 2002; Tokic, 2010; Tokic & Palm, 2011). For instance, such an adaptive policy has been introduced by Tokic and Palm (2011) and is called "Value-difference based exploration combined with softmax action selection" (VDBE-Softmax). This model combines both the ε-greedy and softmax strategy to describe exploratory behavior: While the ε parameter determines the probability to explore in each trial, the softmax policy determines which option to select for exploration. Crucially, the ε parameter in this model depends on the current state of the environment, such that the extent of exploration increases when the knowledge about the environment is uncertain, as indicated by fluctuating values during learning. Tokic and Palm (2011) showed that the VDBE-Softmax model outperforms both the ε-greedy rule and softmax rule in a simulation experiment based on the multi-armed bandit problem.

As demonstrated by this overview, a large number of cognitive models with varying levels of complexity have been developed to empirically describe explore/exploit behavior and to find optimal decision strategies for explore/exploit problems. However, research has so far yielded little or ambiguous evidence as to which of these models best describes human choice behavior and its underlying processes (e.g. Cogliati Dezza et al., 2017; Daw et al., 2006; Speekenbrink & Konstantinidis, 2015; Vermorel & Mohri, 2005). One aspect that further complicates the comparison of these different modeling approaches is that explore/exploit problems come in different forms and operationalizations. Even when limited to the bandit paradigm, the exact circumstances under which explore/exploit behavior is examined may hugely impact the modeling results. Such aspects are, for instance, if the model is applied to a stationary or restless bandit task, if rewards are binary or continuous, if the task horizon is short or long, how outcome values and uncertainties are learned,

as well as the degree of volatility and complexity of the environment (see Behrens, Woolrich, Walton, & Rushworth, 2007; Knox et al., 2012; Payzan-LeNestour & Bossaerts, 2011; Speekenbrink & Konstantinidis, 2015). In the end, more research is needed to further evaluate these different modeling approaches and to resolve the question how explore/exploit behavior can best be described under different circumstances.

1.1.4 Brain regions involved in explore/exploit behavior

The brain regions involved in explore/exploit behavior have been investigated by a number of different studies so far, most of which used model-based functional magnetic resonance imaging (fMRI; see 2.8.1). In the following, results on exploratory choices are reviewed first, while results on exploitative choices will be considered thereafter. The first fMRI experiment that examined the neural correlates of human explore/exploit behavior in a restless four-armed bandit task was the study of Daw et al. (2006). In this study, subjects' choice behavior was first modeled with a Bayesian learner plus softmax rule to obtain trial-by-trial estimates for the expected reward of each bandit. Choices were then classified as either exploitative, i.e. following the highest expected reward, or exploratory, i.e. choosing one of the remaining options. Based on this classification, differential brain activations for exploration and exploitation were analyzed. Exploratory choices were found to specifically activate the right frontopolar cortex (FPC) and, to a lesser extent, the left frontopolar cortex. The FPC is the most anterior part of the prefrontal cortex and is known to be involved in high-level behavioral control (Braver & Bongiolatti, 2002; Christoff & Gabrieli, 2000; Koechlin & Hyafil, 2007; Ramnani & Owen, 2004; Tsujimoto, Genovesio, & Wise, 2011). Hence, its activation in exploratory trials has been proposed to reflect a top-down control mechanism that overrides value-driven choice tendencies to facilitate behavioral switching between an exploitative and exploratory mode (Daw et al., 2006; see also Mansouri, Koechlin, Rosa, & Buckley, 2017). Additionally, greater activation in exploratory trials was also found bilaterally in the intraparietal sulcus (IPS), an area which has repeatedly been implicated in reward-based decision making (Dorris & Glimcher, 2004; Hunt et al., 2012; McClure, Laibson, Loewenstein, & Cohen, 2004; Platt & Glimcher, 1999; Sugrue, Corrado, & Newsome, 2004; Tanaka et al., 2004) and in serving as an interface between perceptive and motor systems for planning and controlling hand and eye movements (Andersen & Buneo, 2002; Buneo & Andersen, 2006; Culham & Valyear, 2006; Gottlieb, 2007; Grefkes & Fink, 2005). Accordingly, Daw et al. (2006) suggested that this region might act as an interface between frontal areas, in which decision variables relevant for exploratory choices are calculated, and motor areas, in which behavioral responses like button presses are generated (see also Rathelot, Dum, & Strick, 2017). Other human fMRI studies have later replicated these findings, all showing an increased bilateral activation in the FPC and IPS during exploratory compared to exploitative decisions in the restless bandit task (Addicott, Pearson, Froeliger, Platt, & McClernon, 2014; Laureiro-Martínez et al., 2014; Laureiro-Martínez, Brusoni, Canessa, & Zollo, 2015). Moreover, further neuroimaging studies provided evidence that the FPC indeed tracks decision variables relevant for exploratory choices, such as the reward probability and reward uncertainty of

alternative (unchosen) choice options (Badre, Doll, Long, & Frank, 2012; Boorman, Behrens, & Rushworth, 2011; Boorman, Behrens, Woolrich, & Rushworth, 2009; Cavanagh, Figueroa, Cohen, & Frank, 2012). These studies also found that the FPC tracks the reward uncertainty of alternative choice options especially in subjects who use an uncertainty-guided strategy for exploration (see Badre et al., 2012; Cavanagh et al., 2012). Furthermore, Boorman et al. (2009) showed that the pattern of functional connectivity between the FPC and IPS changes immediately before a voluntary switch to an alternative choice option takes place, suggesting that the FPC engages parietal areas to implement a behavioral switch when it has accumulated sufficient evidence to support such a decision. Taken together, the reported findings support the notion of a frontoparietal top-down mechanism, in which the FPC tracks information relevant for exploratory decisions and functionally interacts with intraparietal regions for the implementation of exploratory actions.

Given the consistent finding of an increased FPC activation during exploration, a number of studies have investigated the causal role of the FPC in exploratory decision making, either by use of the lesion method (Kovach et al., 2012; Mansouri, Buckley, Mahboubi, & Tanaka, 2015) or with brain stimulation techniques (Raja Beharelle, Anjali, Polanía, Hare, & Ruff, 2015; van Holstein, Froböse, O'Shea, Aarts, & Cools, 2018; Zajkowski et al., 2017). For instance, Kovach et al. (2012) investigated exploratory choice behavior of eight patients with FPC lesions in comparison to healthy controls, using the same restless four-armed bandit task as Daw et al. (2006). Surprisingly, the FPC-lesioned patients showed no general impairment in overall task performance or exploratory switching. However, a model-based analysis of choice behavior revealed that the patient group was selectively impaired in the ability to extrapolate short-term reward trends and to use these extrapolations to guide future choices, such as switching away from an exploited bandit when its reward rate suddenly drops. It was hypothesized from these results that the increased FPC activation during exploration, as observed in previous studies (see above), might reflect its role in tracking reward trends in dynamic environments for guiding behavior. Another FPC lesion study in monkeys (Mansouri et al., 2015) investigated the causal role of the FPC in cognitive flexibility, i.e. the cognitive ability to adapt to changing task demands, as assessed with the Wisconsin Card Sorting Test (WCST) and related tasks. While the FPC-lesioned animals were not impaired in their general ability to follow rule switches, they stayed more focused than control monkeys in exploiting the current task under distractions like an intervening task or an unexpected reward. The authors concluded from these findings that the FPC might play a key role in redistributing cognitive resources from the current task to alternative sources of reward in order to explore new choice opportunities. Recently, also non-invasive brain stimulation techniques have been applied to human subjects in order to investigate the causal role of the FPC in exploratory behavior. One study used transcranial direct current stimulation (tDCS) over the right FPC and found that a selective upregulation of the FPC (via anodal tDCS) leads to more exploratory behavior, while a selective downregulation of the FPC (via cathodal tDCS) results in more exploitative behavior in a three-armed bandit task (Raja Beharelle et al., 2015). Moreover, a model-based analysis of the data showed that

FPC downregulation increased the focus on the bandit with the highest expected reward, whereas choice behavior under FPC upregulation was less influenced by expected rewards and more driven by recent negative reward prediction errors. These results suggest that the FPC is part of a neural mechanism that triggers exploration after encountering a surprisingly low choice outcome in order to search for alternative courses of action, in line with the above findings on FPC-lesioned patients (Kovach et al., 2012). Another study, which applied transcranial magnetic stimulation (TMS), showed that inhibition of the right FPC leads to a selective reduction in directed but not random exploration as assessed with the horizon task (Zajkowski et al., 2017). This finding further supports the view that human exploration is not a unitary but a dual process based on both a directed and an undirected strategy (see Wilson et al., 2014), and shows that both types of exploration rely on (at least partly) dissociable neural systems. Recently, another human TMS study (van Holstein et al., 2018) showed that stimulation of the FPC leads to a decrease in reward-related striatal activity as measured by fMRI. This finding provides further evidence that the FPC exhibits top-down control over striatal reward processing, which might contribute to its role in overriding exploitative tendencies to facilitate behavioral switching from exploitation to exploration. Taken together, the results of lesion and brain stimulation studies strongly suggest that the FPC is causally involved in promoting exploration and further support the view of a top-down control mechanism for exploratory decisions.

Aside from the FPC and IPS, there are additional brain regions that have repeatedly been shown to exhibit greater activation during exploration compared to exploitation. Two of these regions are the insula and the anterior cingulate cortex (ACC; Addicott et al., 2014; Blanchard & Gershman, 2018; Laureiro-Martínez et al., 2014, 2015). Consistent with their activation during exploration, both these regions have been shown to be activated during risky decision making and have been implicated in encoding reward uncertainty or risk (Christopoulos, Tobler, Bossaerts, Dolan, & Schultz, 2009; Critchley, Mathiast, & Dolan, 2001; Fitzgerald, Seymour, Bach, & Dolan, 2010; Fukunaga, Purcell, & Brown, 2018; Huettel, Song, & McCarthy, 2005; Preuschoff, Bossaerts, & Quartz, 2006; Preuschoff, Quartz, & Bossaerts, 2008; Rudorf, Preuschoff, & Weber, 2012; see also reviews by Bach & Dolan, 2012; Dreher, 2013; Singer, Critchley, & Preuschoff, 2009). Moreover, both regions have also been implicated in emotional processing and in mediating the effects of emotional arousal on decision making (Bush, Luu, & Posner, 2000; Craig, 2002, 2009; Critchley, 2005; Xue, Lu, Levin, & Bechara, 2010). The anterior insula (AI), in particular, is considered to play a key role in the conscious experience (i.e. the feeling) of an emotion (Craig, 2002, 2009; Damasio & Carvalho, 2013; Singer et al., 2009), and its activation in exploratory trials has been proposed to reflect the experience of anxiety associated with choosing an option with a highly uncertain outcome (Laureiro-Martínez et al., 2015). Furthermore, the AI and dorsal ACC (dACC) are considered to form a "salience network" for the detection of behaviorally relevant stimuli in order to guide attention and actions towards these stimuli (Menon, 2015; Menon & Uddin, 2010; Uddin, 2015). Accordingly, it has been proposed that these regions might subserve attentional and behavioral switching from an exploitative to an exploratory mode (Laureiro-Martínez

et al., 2015; see below). In line with this view, the dACC has been ascribed the function of promoting behavioral switching in foraging decisions (Kolling, Behrens, Mars, & Rushworth, 2012; Kolling, Behrens, Wittmann, & Rushworth, 2016; Rushworth, Kolling, Sallet, & Mars, 2012). Specifically, it has been suggested that the dACC may represent "the value of switching to a course of action alternative to that which is taken or is the default" (Kolling et al., 2012, p. 97; although see Shenhav, Straccia, Cohen, & Botvinick, 2014), consistent with its proposed role in switching between exploitation and exploration. Aside from the insula and ACC, exploratory choices were also shown to be associated with higher activity in the cerebellum, thalamus, supplementary motor area (SMA), and brain stem (Addicott et al., 2014; Daw et al., 2006, supplement; Laureiro-Martínez et al., 2014, 2015). Concerning the increased brain stem activation during exploration, it has been proposed that these signals might originate from the locus coeruleus (LC; Laureiro-Martínez et al., 2014, 2015), although temporal resolution in these fMRI studies was too low to unambiguously confirm this (see Addicott et al., 2017). However, an increased LC activation during exploration would support the view that the LC norepinephrine system is also playing an important role in regulating the trade-off between exploration and exploitation (see 1.1.5; e.g. Aston-Jones & Cohen, 2005; Cohen et al., 2007).

In contrast to exploration, research on the brain regions involved in exploitation has yielded more mixed results so far. For instance, Daw et al. (2006) did not find any brain regions showing significantly greater activity during exploitative compared to exploratory trials. However, they reported that both the orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC) exhibit "activity characteristic of an involvement in value-based exploitative decision making" (Daw et al., 2006, p. 876). More specifically, they found activity in the medial OFC to correlate with the magnitude of the obtained reward, and activity in the vmPFC and medial/lateral OFC to correlate with the choice probability of the chosen option. Note that this choice probability reflects the expected reward of the chosen option relative to the unchosen ones, according to the standard softmax model they applied (see also 2.7.2). Later studies supported these findings by showing that both the vmPFC and OFC exhibit significantly greater activity during exploitative compared to exploratory trials (see below; Laureiro-Martínez et al., 2014, 2015). Moreover, the observation of a reward- and exploitation-related activity in the vmPFC and OFC converges well with a large body of research implicating these regions in encoding the subjective value of both primary and secondary rewards during decision making and outcome delivery (see reviews by Grabenhorst & Rolls, 2011; Kringelbach & Rolls, 2004; O'Doherty, 2004; Rushworth et al., 2012). Together with the ventral striatum, these prefrontal regions are thought to form the core of a "valuation system" in the human brain, which codes a domain-general subjective value signal and plays a key role in guiding reward-based decision making (Bartra, McGuire, & Kable, 2013; Fellows, 2011; Kable & Glimcher, 2009; Levy & Glimcher, 2012; O'Doherty, 2011; Peters & Büchel, 2010). In contrast to Daw et al. (2006), later human fMRI studies with larger sample sizes have revealed several brain regions that show significantly greater activation during exploitative compared to exploratory choices in the restless bandit task. For example, Laureiro-Martínez et al. (2014) showed that

exploitative choices elicited stronger activation in a number of distributed brain regions, including the vmPFC and OFC, as well as the bilateral hippocampus, ACC, middle temporal gyri, and left posterior cingulate cortex (PCC). A second study from the same group (Laureiro-Martínez et al., 2015) largely replicated these findings, also showing exploitation-specific activity in regions of the vmPFC and OFC, the bilateral hippocampus, the left ACC and middle temporal gyrus, as well as in the bilateral superior temporal gyri and left PCC/precuneus. Note, however, that the latter study used a behavioral rather than a computational definition of explore/exploit decisions, in which exploitation was defined as staying with the current option and exploration as switching to an alternative option. A third human fMRI study (Addicott et al., 2014), on the other hand, reported greater activation during exploitation compared to exploration only in the left angular gyrus and bilateral temporal lobes, including the superior/middle temporal gyri and planum temporale, but not in prefrontal regions like the vmPFC or OFC. The failure of this study to detect prefrontal activation during exploitation might be (partly) due to its relatively low sample size (n=24), which was less than half of the sample sizes used in the two studies by Laureiro-Martínez et al. (2014, 2015). For a discussion on the potential functional roles of these brain regions in exploitative decision making, the reader is also referred to section 6.3.1. Taken together, although research on the neural correlates of exploitation yielded partly mixed results, most of these studies provided evidence for an involvement of prefrontal areas, especially the vmPFC and OFC, in reward coding and exploitative decision making.

Aside from mapping the distinct neural signatures of exploratory and exploitative choices, research has also provided first insights into the neural mechanism which may underlie switching between exploitation and exploration. For example, a model-based fMRI study by Boorman et al. (2009) investigated the neural precursors to behavioral switching in a restless two-armed bandit task in humans. Notably, this study found that the vmPFC encodes the value of the chosen relative to the unchosen action, while the FPC tracks the reward probability of the unchosen action relative to the chosen one. Moreover, FPC activity during the intertrial interval was found to predict behavioral switching to the alternative option on the following trial within and between subjects, and to functionally interact with the IPS immediately before such a switch takes place (see above). Hence, these results suggest that the vmPFC and FPC play key complementary roles in human decision making (see also Domenech & Koechlin, 2015) and corroborate well with the above findings, showing greater vmPFC activation during exploitation and greater FPC and IPS activation during exploration. Moreover, the results of various neuroimaging studies have recently been integrated by Laureiro-Martínez et al. (2015) into a proposed neural mechanism for behavioral switching between exploitation and exploration. In short, they suggest that the values of foregone options, as represented in the FPC (Boorman et al., 2009, 2011; Mansouri et al., 2017), are continuously compared with the value of the current choice, as represented in medial prefrontal regions (Boorman et al., 2009; Daw et al., 2006), by the monitoring performance mechanisms implemented in the ACC (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). If activity in the FPC exceeds vmPFC activity, i.e. if the accumulated evidence

favors the alternative over the current choice option, this triggers an attentional disengagement from the current choice in the IPS (see Boorman et al., 2009; Laureiro-Martínez et al., 2015). Hence, switching between exploitation and exploration appears to require the interplay of several brain regions subserving different functions, including the evaluation of competing choice options, performance monitoring, attentional/behavioral control, and action implementation.

Finally, it should be noted that while explore/exploit decisions have mostly been studied within the multi-armed bandit paradigm, this paradigm has recently been criticized to not clearly disentangle exploration from exploitation, making it therefore problematic to unambiguously identify their neural correlates (see Blanchard & Gershman, 2018). A first point of criticism relates to the fact that exploration in the bandit task not only yields information but also reward, while exploitation not only yields reward but also information, making it difficult to conceptually differentiate between both types of decisions as either purely reward-maximizing or purely information-seeking. A further point of criticism has been that the classification of choices as exploitative or exploratory is based on subjective reward estimates derived from a cognitive model, which in turn depends on prior assumptions about a subject's decision strategy, offering no theory-neutral way to distinguish between both types of decisions. To overcome these drawbacks of the bandit task, the "observe or bet" task (Tversky & Edwards, 1966) has been proposed by Blanchard and Gershman (2018) as an alternative to clearly distinguish between explore/exploit decisions and their neural signatures. In this task, subjects need to decide to either observe the outcome of an option without actually gaining it ("pure exploration") or to bet on an option and gain its reward without receiving direct feedback about the rewarded value during the task ("pure exploitation"). Based on this task, Blanchard and Gershman (2018) investigated the neural correlates of pure explore/exploit decisions in humans by comparing brain activation patterns between observe and bet trials using fMRI. In line with findings based on the bandit paradigm, they found that pure exploration is associated with greater activation in the dACC, insula, and thalamus, while pure exploitation is associated with greater vmPFC activation, although the latter result did not survive multiple comparison correction. Yet, surprisingly, they did not find higher activity during pure exploration in the FPC, which was included in their region of interest analysis, nor in the IPS within a whole-brain analysis, suggesting that this task indeed measures (partly) different choice processes than the restless bandit task. While the "observe or bet" task offers an interesting model-free approach to study explore/exploit behavior, it also brings the limitation of a task switching confound due to the dual nature of the task (for further details see Blanchard & Gershman, 2018). Moreover, it is unclear how well the concept of a binary distinction between pure exploration and pure exploitation actually represents real-world explore/exploit decisions, which most often come with a coupling of reward and information (e.g. in foraging), as implemented by the bandit task. Still, research on explore/exploit behavior might benefit from the application of such different experimental paradigms that allow to assess different aspects of this complex behavior and their underlying neural processes.

In conclusion, neuroscientific research has already made some progress towards mapping the brain regions involved in exploration and exploitation. Furthermore, the knowledge gained from these studies has also been used for the development of neural network models, which seek to explain how these different brain areas functionally interact to initiate switching between both decision strategies (e.g. Cohen et al., 2007; Laureiro-Martínez et al., 2014, 2015; McClure, Gilzenrat, & Cohen, 2006). Still, more research is needed to test and refine these neuromechanistic accounts, e.g. by applying model-based fMRI and non-invasive brain stimulation techniques (see Parkin, Ekhtiari, & Walsh, 2015) to further elucidate the neural computations underlying explore/exploit decisions and their causal role in this trade-off.

1.1.5 Neurochemical systems involved in explore/exploit behavior

This section will give a brief overview of the main neurochemical systems (i.e. neurotransmitters and neuromodulators) involved in explore/exploit behavior, including dopamine, norepinephrine, and acetylcholine.

Dopamine (DA) is one of the most studied neurochemicals in research on reward-based decision making and explore/exploit behavior, whereby different aspects of DA signaling are believed to subserve different behavioral functions (e.g. phasic vs. tonic DA, see 1.2.2). First of all, it is by now well established that phasic DA is tightly involved in the process of reinforcement learning (RL) by encoding a reward prediction error (RPE) signal (reviewed by Glimcher, 2011; Schultz, 2016; Schultz, Stauffer, & Lak, 2017). This RPE signal reflects differences between received and expected outcomes and serves as a "teaching signal" according to temporal difference models of RL (see 1.1.3; e.g. Sutton & Barto, 2018). Neuroimaging studies in animals and humans provide strong evidence that midbrain DA neurons encode the RPE signal by responding with short phasic bursts (or pauses) of firing to discrete events involving errors in reward prediction on a millisecond timescale (Hart, Rutledge, Glimcher, & Phillips, 2014; Ljungberg, Apicella, & Schultz, 1992; Schultz, Apicella, & Ljungberg, 1993; Schultz, Dayan, & Montague, 1997; Waelti, Dickinson, & Schultz, 2001; Zaghloul et al., 2009). In addition, more recent experiments in animals and humans yielded causal evidence for the notion that this phasic DA activity actually drives reward learning and reward-seeking behavior (Adamantidis et al., 2011; Chang et al., 2018; Kim et al., 2012; Ramayya, Misra, Baltuch, & Kahana, 2014; Steinberg et al., 2013; Tsai et al., 2009; Witten et al., 2011; Zweifel et al., 2009; see also reviews by Schultz et al., 2017; Steinberg & Janak, 2013). Hence, given these findings, phasic DA signaling appears to play a causal role in driving exploitative behavior. Aside from that, other studies have also focused more specifically on DA's role in regulating the trade-off between exploitation and exploration – a topic which is considered in more detail in a later section of this introduction (see 1.3). To shortly summarize, evidence from animal and neural network studies suggest that tonic DA may be involved in regulating the trade-off between exploitation and random exploration by controlling the degree to which behavioral choices are based on previously learned rewards (Beeler, Daw, Frazier, & Zhuang, 2010; Beeler, Frazier, & Zhuang, 2012;

Humphries, Khamassi, & Gurney, 2012). Moreover, also human studies have provided first evidence for a DA involvement in controlling explore/exploit behavior, which is mainly based on findings from genetic variation studies in healthy subjects (Blanco et al., 2015; Frank et al., 2009; Kayser, Mitchell, Weinstein, & Frank, 2015) or from studies on altered explore/exploit behavior in patients with DA-related disorders, such as Parkinson's disease or schizophrenia (e.g. Moustafa et al., 2008; Strauss et al., 2011). Finally, indirect evidence also stems from animal and human studies showing that DA modulates potential subcomponents of explore/exploit behavior, such as risky decision making (Kohno et al., 2016; Lancaster, Linden, & Heerey, 2012; Sherman & Wilson, 2016; St Onge, Abhari, & Floresco, 2011) and cognitive or behavioral flexibility (Beeler et al., 2014; Cools & D'Esposito, 2011; Floresco, 2013).

Aside from DA, also the locus coeruleus norepinephrine (LC-NE) system is considered to play an important role in the regulation of explore/exploit behavior (Aston-Jones & Cohen, 2005; Cohen et al., 2007; McClure et al., 2006). The locus coeruleus is a brain stem nucleus containing NE-synthesizing neurons that send widespread projections throughout the whole brain, including the neocortex, hippocampus, cerebellum, and thalamus (Benarroch, 2009). According to the "adaptive gain theory" of LC-NE function (Aston-Jones & Cohen, 2005), which is rooted in monkey neurophysiological research, the trade-off between exploration and exploitation can be explained through two different modes of LC activity: phasic and tonic. A phasic LC mode is characterized by short bursts of LC activity in response to task-related stimuli, which facilitates engagement in the current task and the optimization of task performance, i.e. exploitative behavior. In contrast, a tonic LC mode is characterized by a higher baseline LC activity and reduced bursts, which promotes disengagement from the current task and the search for alternative options, i.e. exploratory behavior. According to the theory, transitions between the phasic and tonic LC mode are triggered by changes in task-related utility, which is signaled to the LC via direct inputs from the ACC and OFC, where task performance and utility are monitored. Specifically, high values of long-term task utility drive the phasic (exploitative) mode, while low values trigger the tonic (exploratory) mode. Based on the assumptions of the adaptive gain theory, several human studies have been performed to investigate the role of the LC-NE system in explore/exploit behavior. Most of these studies applied pupillometry, the measuring of pupil size, as an indirect marker of LC activity during task performance (see Aston-Jones & Cohen, 2005; Murphy, Robertson, Balsters, & O'Connell, 2011). For instance, Jepma and Nieuwenhuis (2011) assessed subjects' pupil size while they made explore/exploit decisions in a restless four-armed bandit task and found that exploratory choices were preceded by a larger baseline pupil diameter (reflective of a higher tonic LC activity) than exploitative choices. Moreover, they showed that changes in pupil size during explore/exploit transitions correlated with changes in task utility, and that subjects with a larger baseline pupil diameter showed an overall higher tendency to explore. All these findings are in line with the adaptive gain theory, supporting the view that utility-related changes of LC activity play an important role in the regulation of explore/exploit behavior. In addition, also other studies reported

increased baseline pupil diameter during exploratory behavior using different behavioral paradigms, including a target detection task (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010), an analogical reasoning task (Hayes & Petrov, 2016), and attentional set shifting tasks (Pajkossy, Szőllősi, Demeter, & Racsmány, 2017). On the other hand, two pharmacological studies reported findings not supporting the adaptive gain theory. The first study (Jepma, Te Beek, Wagenmakers, van Gerven, & Nieuwenhuis, 2010) did not find any differences in exploratory behavior under reboxetine, a selective NE reuptake inhibitor that increases tonic NE levels, although changes in other behavioral parameters indicated the general effectiveness of the drug treatment. In the second study (Warren et al., 2017), administration of atomoxetine, another selective NE reuptake inhibitor, was found to reduce rather than increase random exploration in the horizon task, while leaving directed exploration unaffected. In both studies, the adaptive gain theory would instead have predicted an increase in exploration driven by a drug-induced increase in tonic NE levels. Given these inconsistent results, further research will be necessary to clarify the role of the LC-NE system in human explore/exploit behavior.

Recently, also the role of acetylcholine (ACh) in explore/exploit behavior has been investigated in a transgenic mice experiment (Naudé et al., 2016). The study found that mice lacking the β2* subunit of the nicotinic ACh receptor (nAChR) showed less uncertainty-driven exploration than wild-type controls in a mice-adapted three-armed bandit task. The nAChR is a receptor expressed in the ventral tegmental area (VTA), where it is thought to influence DA transmission and thereby value-based decision making through a yet unknown mechanism (see Naudé et al., 2016). Furthermore, the study showed that re-expression of the β2* subunit in the transgenic mice restored spontaneous activity of DA neurons in the VTA as well as uncertainty-driven exploration. The authors conclude from these findings that the nicotinic ACh receptor in the VTA is involved in translating uncertainty into motivational value for driving exploratory decisions, which might also explain altered risk-taking and exploratory behavior observed in nicotine addiction (e.g. Addicott, Pearson, Wilson, Platt, & McClernon, 2013; Lejuez, Aklin, Bornovalova, & Moolchan, 2005).

Finally, when studying the role of different neuromodulatory systems in explore/exploit behavior, it is important to note that these systems do not function separately in the brain, but tightly interact with each other. The study of Naudé et al. (2016) provides an example for the interaction between the DA and ACh system, but also interactions between DA and NE have often been claimed as important for the regulation of explore/exploit behavior (Aston-Jones & Cohen, 2005; Cohen et al., 2007; McClure et al., 2006). For example, the adaptive gain theory (Aston-Jones & Cohen, 2005) assumes that the LC-NE system synergistically interacts with DA-dependent reward learning for controlling explore/exploit behavior: During the tonic LC mode (exploration), when utility is low, the DA system drives reinforcement learning to discover new sources of reward and to strengthen behavior towards these rewards. This reinforced behavior, in turn, leads to an increase in utility and a transition of LC activity from the tonic into the phasic mode (exploitation). During exploitation, the rewarded behavior is maintained by both the phasic LC activity and DA-dependent reinforcement, until utility declines and

the LC changes back into the tonic mode to promote exploration of new resources. Moreover, it has also been proposed that NE interacts with the ACh system to signal different types of uncertainty for decision making (Yu & Dayan, 2005). Specifically, it was suggested that ACh levels encode "expected uncertainty", i.e. the uncertainty which can be predicted from prior experience (e.g. probabilistic rewards), whereas NE levels signal "unexpected uncertainty", i.e. unforeseen fluctuations of decision outcomes that might call for an update of beliefs and a change of behavioral strategy. As proposed by Yu and Dayan (2005), the direct comparison of these two uncertainty signals in the brain might provide a computationally tractable algorithm for determining when to revise expectations, which is likely to be involved in controlling explore/exploit behavior. For instance, when the uncertainty starts to exceed the expected degree signaled by ACh, then this rise in unexpected uncertainty, signaled by NE, might drive exploratory behavior (see Cohen et al., 2007). This assumption is also in line with the adaptive gain theory, according to which an increase in tonic LC activity promotes exploration (Aston-Jones & Cohen, 2005).

Taken together, the current research literature provides strong evidence that different neurochemical systems – including DA, NE, and ACh – and their interactions play an important role in the regulation of the explore/exploit trade-off. The next part of this introduction will first broadly introduce the DA brain system and its basic characteristics in general, before the current state of research on the role of DA in explore/exploit behavior is reconsidered in more detail.

1.2 The dopaminergic brain system

1.2.1 Dopamine and dopaminergic pathways in the brain

Dopamine (DA) is an organic compound of the catecholamine group that exhibits several important functions in the body and brain. Chemically, dopamine (3,4-dihydroxyphenethylamine) is characterized by a benzene ring with two hydroxyl groups and an amine side chain. It is synthesized in the organism from its precursor L-dopa (L-3,4-dihydroxyphenylalanine) by enzymatic decarboxylation. In turn, DA itself represents a precursor molecule for the synthesis of the catecholamine neurotransmitters norepinephrine and epinephrine (Meiser, Weindl, & Hiller, 2013). In the brain of animals and humans, DA acts as both a synaptic neurotransmitter and neuromodulator (Doya, 2002; Katz & Calin-Jageman, 2009; Richerson, Aston-Jones, & Saper, 2013; Schultz, 2007). As a neurotransmitter, DA is involved in the fast, time-specific signal transduction between a pre- and postsynaptic neuron via impulse-dependent synaptic DA release (see reviews by Grace, Lodge, & Buffalari, 2009; Schultz, 2007; Schultz et al., 2017). As a neuromodulator, DA diffuses through the brain tissue and can thereby exert spatially distributed, temporally extended effects on several neurons distant from the site of release, e.g. by modulating excitability and synaptic strength at these targets (Marder, 2012; Marder & Thirumalai, 2002; Nadim & Bucher, 2014; Richerson et al., 2013; Schultz, 2007). The majority of DA neurons are localized in the mesencephalon (midbrain), mainly in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA; Arias-Carrión, Stamelou, Murillo-Rodríguez, Menéndez-González, &

Pöppel, 2010). Although these DA neurons make up less than 1% of the neuronal brain cell population (Arias-Carrión & Pöppel, 2007; Björklund & Dunnett, 2007), they affect diverse brain functions by sending numerous and widespread projections to different subcortical and cortical regions.

Three main DA pathways, or DA systems, are distinguished in the brain (see Figure 1; Arias-Carrión et al., 2010; Ayano, 2016). First, the nigrostriatal pathway, projecting from the SNc to the dorsal striatum, i.e. the caudate nucleus and putamen. This nigrostriatal system is part of the basal ganglia circuit and plays a key role in voluntary movement control and motor skill learning (Bissonette & Roesch, 2015; Korchounov, Meyer, & Krasnianski, 2010; Obeso et al., 2008; Smith & Villalba, 2008). Second, the mesolimbic pathway, originating in the VTA and projecting to the ventral striatum, i.e. the nucleus accumbens and olfactory tubercle, but also to the amygdala, hippocampus, and septum. This pathway is considered to be crucially involved in reinforcement learning and motivated behaviors (Arias-Carrión et al., 2010; Berridge, 2012; Berridge & Kringelbach, 2015; Bissonette & Roesch, 2015). Third, the mesocortical pathway, sending projections from the VTA to cortical regions, e.g. the prefrontal, cingulate, and perirhinal cortex. This pathway is regarded to play an important role in executive functions, including working memory and cognitive or behavioral flexibility (Cools & D'Esposito, 2011; Floresco, 2013; Floresco & Magyar, 2006; Klanker, Feenstra, & Denys, 2013; Leh, Petrides, & Strafella, 2010). Because the mesolimbic and mesocortical pathway are tightly connected, they are often together referred to as the mesocorticolimbic system (see Arias-Carrión et al., 2010; Wise, 2005).

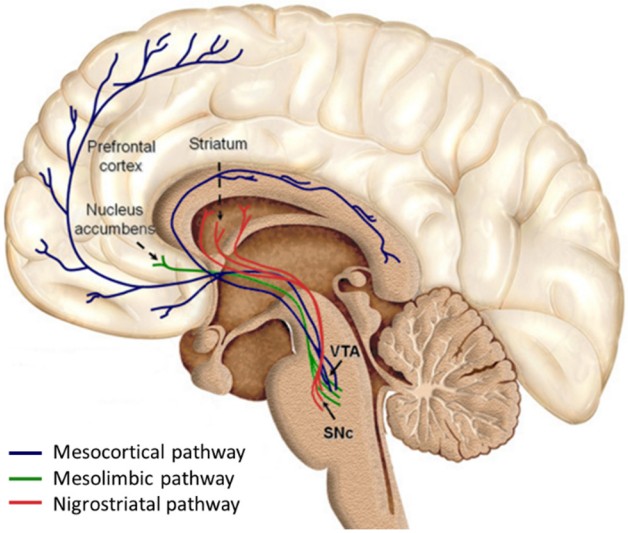


Figure 1. Main dopaminergic pathways in the human brain. SNc: substantia nigra pars compacta; VTA: ventral tegmental area. Adapted from Arias-Carrión et al. (2010).

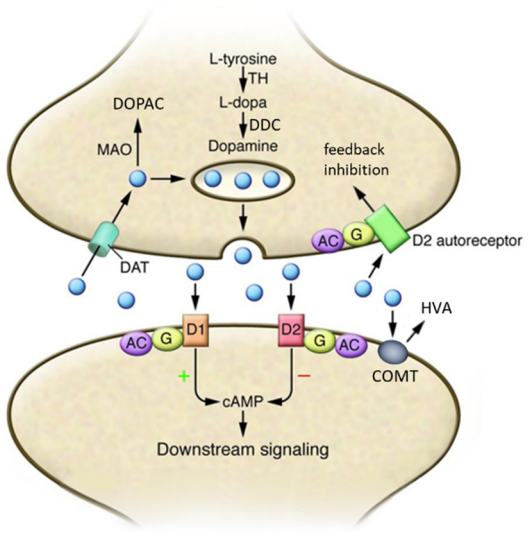


Figure 2. Dopaminergic synapse. See text for further explanations and abbreviations. Adapted from Blackstone (2009).

1.2.2 DA neurotransmission

Dopaminergic neurotransmission involves several processes and molecules (see Figure 2). In the cytosol of catecholamine-producing neurons, DA is synthesized from its precursor L-dopa by the enzyme aromatic L-amino acid decarboxylase (AADC), also known as DOPA decarboxylase (DDC).

L-dopa itself is produced from the amino acid L-tyrosine by the enzyme tyrosine hydroxylase (TH), which catalyzes the initial and rate-limiting step in the biosynthesis of DA (see Hälbig & Koller, 2007; Meiser et al., 2013). After cytosolic synthesis, DA is packaged and stored in synaptic vesicles located within the presynaptic terminals of DA neurons. When an action potential reaches the axon terminal, it triggers a series of events causing DA to be released into the synaptic cleft by a process called exocytosis (Westerink, 2006). Upon synaptic release, DA diffuses across the synaptic cleft to the postsynaptic membrane, where it binds to and activates DA receptors. Importantly, there are several subtypes of DA receptors that exhibit different cellular functions, adding to the complexity of the DA system and its physiological functions (for details see Beaulieu & Gainetdinov, 2011; Neve, 2010; Neve, Seamans, & Trantham-Davidson, 2004). In mammals, there are at least five subtypes of DA receptors, labeled D1 to D5, whereby D1 and D2 are the most frequent subtypes in the human brain (Ayano, 2016; Hurd, Suzuki, & Sedvall, 2001; Strange & Neve, 2013). These five subtypes are grouped into two families, the D1-like and D2-like receptor family, which differ in their pharmacological properties and associated signaling pathways (see below). Both DA receptor families belong to the class of G protein-coupled receptors (GPCRs), also called seven-transmembrane (7TM) receptors as they span the cell membrane seven times (Oldham & Hamm, 2008; Pierce, Premont, & Lefkowitz, 2002). The binding of extracellular DA to these receptors triggers an intracellular signaling cascade mediated by a guanine nucleotide-binding protein (G-protein), which is coupled to the cytosolic part of the receptor. This G-protein mediated signaling cascade can have different downstream effects on the cell, depending on the type of G-protein that is activated by the receptor (see Beaulieu & Gainetdinov, 2011; Romanelli, Williams, & Neve, 2010). D1-like receptors (D1 and D5) are coupled to $G\alpha_{s/olf}$ proteins, which stimulate the regulatory enzyme adenylate cyclase (AC) in its production of the second messenger cyclic adenosine monophosphate (cAMP). In contrast, D2-like receptors (D2, D3, D4) are coupled to $G\alpha_{i/o}$ proteins, which inhibit adenylate cyclase and cAMP production. This increase or decrease in cellular cAMP levels can have different downstream effects on a wide array of cellular substrates, including ion channels, enzymes, and transcription factors (Neve et al., 2004; Tritsch & Sabatini, 2012). Hence, the same ligand, DA, can trigger very different cellular responses in a target neuron, depending on the subtype of DA receptor and the intracellular signaling pathways it activates. In addition to postsynaptic DA receptors, there are also presynaptic DA autoreceptors of the D2 or D3 subtype residing in the cell membrane of an axon terminal (Ford, 2014; Schmitz, Benoit-Marand, Gonon, & Sulzer, 2003; Wolf & Roth, 1990). These autoreceptors play an important role in regulating DA transmission via a negative feedback mechanism (feedback inhibition), which controls different presynaptic cell processes in response to extracellular DA levels, such as firing patterns, DA synthesis, DA release, and DA uptake (de Mei, Ramos, Iitaka, & Borrelli, 2009; Ford, 2014; Sulzer, Cragg, & Rice, 2016). Furthermore, extracellular DA can diffuse away from the synaptic cleft into the extrasynaptic space of the surrounding brain tissue and act on more distant targets as a neuromodulator (see above; e.g. Nadim & Bucher, 2014; Richerson et al., 2013; Schultz, 2007).

DA neurotransmission can be terminated by different processes, either by rapid reuptake of DA into the presynaptic neuron or by metabolic degradation (see Figure 2). Reuptake is mainly mediated by the dopamine (active) transporter, short DAT, which spans the presynaptic cell membrane and actively pumps most of the released DA back into the cytosol, where it is repacked into storage vesicles or degraded (see Sotnikova, Beaulieu, Gainetdinov, & Caron, 2006; Torres, Gainetdinov, & Caron, 2003; Westerink, 2006). DA inactivation by metabolic degradation, on the other hand, primarily involves the enzymes catechol-O-methyltransferase (COMT) and monoamine oxidase (MAO), resulting in the main degradation products homovanillic acid (HVA) and 3,4-dihydroxyphenylacetic acid (DOPAC; Hälbig & Koller, 2007; Meiser et al., 2013). While MAO is located in the outer mitochondrial membrane of neurons and glial cells (Meiser et al., 2013; Westlund, Denney, Rose, & Abell, 1988), COMT is mainly found in glial cells and only some types of neurons, where it is predominantly expressed in its membrane-bound isoform (Chen et al., 2011; Männistö & Kaakkola, 1999; Meiser et al., 2013; Myöhänen & Männistö, 2010; Myöhänen, Schendzielorz, & Männistö, 2010). It is still controversial, however, whether this membrane-bound COMT isoform is bound to the plasma membrane or to intracellular membranes and if it acts extra- or intracellularly to degrade DA (see Chen et al., 2011; Myöhänen & Männistö, 2010; Schott et al., 2010). The soluble COMT isoform, on the other hand, is mainly expressed in peripheral tissues and considered to play only a minor role for DA inactivation in the human brain (Chen et al., 2011; Männistö & Kaakkola, 1999; Myöhänen & Männistö, 2010). Importantly, the two mechanisms of DA reuptake and DA degradation differentially contribute to the termination of DA transmission in different brain regions. Fast reuptake via DAT is the primary way of terminating the DA signal in the striatum (Cass & Gerhardt, 1995; Cass, Zahniser, Flach, & Gerhardt, 1993; Giros, Jaber, Jones, Wightman, & Caron, 1996; Shen et al., 2004). However, DA reuptake plays a minor role in regions with low DAT expression, such as the prefrontal cortex (PFC), where degradation by COMT appears to substantially contribute to DA inactivation (Bilder, Volavka, Lachman, & Grace, 2004; Garris & Wightman, 1994; Käenmäki et al., 2010; Sesack, Hawrylak, Matus, Guido, & Levey, 1998; Tunbridge, Bannerman, Sharp, & Harrison, 2004; Yavich, Forsberg, Karayiorgou, Gogos, & Männistö, 2007). As COMT inactivates DA more slowly than DAT, DA signals persist much longer in the PFC, and the released DA is free to diffuse out of the synaptic cleft to have more widespread effects on extrasynaptic sites (see Bilder et al., 2004; Cass & Gerhardt, 1995; Garris & Wightman, 1994; Sesack et al., 1998).

Finally, two different modes of DA neurotransmission can be distinguished, known as the phasic and tonic mode (Grace, 1991; Grace & Bunney, 1984a, 1984b; Hyland, Reynolds, Hay, Perk, & Miller, 2002; Owesson-White et al., 2012; see also reviews by Goto, Otani, & Grace, 2007; Wightman & Robinson, 2002). Phasic DA transmission is caused by a short high-frequency series of action potentials (“burst”), e.g. triggered by a rewarding stimulus or electrical stimulation of DA neurons, which leads to a fast presynaptic DA release. This burst firing rapidly and transiently increases DA concentrations in the synaptic cleft to the micro- to millimolar range, which is sufficient to stimulate postsynaptic DA

receptors for signal transduction (Bilder et al., 2004; Goto et al., 2007; Grace, 1991, 2008). Tonic DA transmission, in contrast, is caused by a low-frequency (∼5 Hz) series of action potentials, which occur spontaneously and irregularly due to the baseline activity of DA neurons. This baseline firing results in a slow DA release that underlies the constant low levels of tonic DA in the extrasynaptic space, which reach concentrations in the tens of nanomolar range (Abercrombie, Keefe, DiFrischia, & Zigmond, 1989; Grace, 1991, 2008; Smith, Olson, & Justice, 1992; Zetterström, Sharp, Marsden, & Ungerstedt, 1983). While this concentration is too low to activate postsynaptic DA receptors for signal transduction, it is sufficient to stimulate presynaptic DA autoreceptors and can thereby act to modulate (i.e. downregulate) phasic DA transmission (Bilder et al., 2004; Floresco, West, Ash, Moore, & Grace, 2003; Grace, 1991). Furthermore, the mechanisms of tonic and phasic DA transmission are considered to differ between brain regions (see Bilder et al., 2004): In the striatum, tonic DA release is thought to result from the slow baseline firing of VTA neurons, as described above, and may further be modulated by glutamatergic afferents from the PFC. In the PFC, however, tonic DA levels are believed to result from the phasic burst firing of VTA neurons and the subsequent diffusion of phasically released DA into the extrasynaptic space. The diffusion of DA out of the synaptic cleft is facilitated in the PFC due to a reduced clearance of synaptic DA by DAT, which is much lower expressed in prefrontal compared to striatal regions (see above). Thereby, phasic burst firing of VTA neurons may contribute to both phasic and tonic DA transmission in the PFC. Finally, the phasic and tonic mode of DA activity are believed to mediate distinct aspects of behavior (see Beeler et al., 2012; Beeler, 2012; Goto et al., 2007; Schultz, 2002, 2007; Zweifel et al., 2009). Phasic DA activity is considered to encode an RPE signal that drives reinforcement learning, mediated by the activity of midbrain DA neurons that respond with short phasic bursts (or pauses) of firing to errors in reward prediction on a millisecond timescale (see 1.1.5; e.g. Glimcher, 2011; Schultz, 2016). Tonic DA activity, on the other hand, is thought to operate across a broader temporal span than phasic DA and is more widely associated with motivational aspects of reward-based behavior, such as the regulation of response vigor (Beierholm et al., 2013; Niv, Daw, Joel, & Dayan, 2007; Rigoli, Chew, Dayan, & Dolan, 2016) and behavioral energy expenditure (see 1.3.1; Beeler et al., 2012; Beeler, 2012).

Taken together, this section provided a brief overview of the different neural pathways, receptors, and cellular mechanisms involved in DA signaling, which contribute to the immense complexity of the DA brain system and the diversity of physiological functions mediated by DA. For a more comprehensive introduction to these topics, the reader is referred to Neve (2010), Beaulieu and Gainetdinov (2011), and Iversen (2010).

1.2.3 DA pharmacology

Dopaminergic drugs modulate the function of the DA brain system by specifically targeting the molecules, enzymes, and processes involved in DA signaling. For example, DA drugs can act as (selective) DA receptor agonists or antagonists, thereby stimulating or blocking the activity of one or

more subtypes of the DA receptor (see Beaulieu & Gainetdinov, 2011; Nichols, 2010; Prante, Dörfler, & Gmeiner, 2010). Other DA drugs target the metabolic pathway of DA, e.g. by inhibiting DA degradation by COMT or MAO (Nissinen & Männistö, 2010; Pisani et al., 2011; Scheggia, Sannino, Scattoni, & Papaleo, 2012; Stahl & Felker, 2008), or by providing a precursor substance for increased DA synthesis like the drug L-dopa (see below). Furthermore, DA drugs can also affect synaptic release or reuptake of DA, for instance by inhibiting DAT or other DA transporter molecules (Gether, Andersen, Larsson, & Schousboe, 2006; Huot, Fox, & Brotchie, 2016). As a result, DA drugs may lead to an overall increase or decrease in DA transmission (the latter also referred to as "antidopaminergic effects"), but mostly affect the function of the DA system in more complex ways (see below; e.g. Cools, 2006; Pryor & Storer, 2013; Ruskin et al., 1999; Stępnicki, Kondej, & Kaczor, 2018; Yael et al., 2013).

Dopaminergic drugs have a broad field of applications. First of all, DA drugs are often developed and used for medical treatment of human diseases known to involve a DA dysfunction, such as Parkinson's disease and schizophrenia. Parkinson's disease (PD) is characterized by a progressive degradation of DA neurons in the basal ganglia, leading to striatal DA depletion and severe motoric symptoms, such as slowed voluntary movement (bradykinesia), tremor, and rigidity (Hornykiewicz, 1966; Kish, Shannak, & Hornykiewicz, 1988; Litvan et al., 2003). Schizophrenia, on the other hand, is considered to involve excessive striatal DA function giving rise to positive symptoms like hallucinations and delusions, and reduced prefrontal DA function resulting in cognitive impairments and negative symptoms like avolition, anhedonia, and alogia (Abi-Dargham, 2004; Davis, Kahn, Ko, & Davidson, 1991; Howes & Kapur, 2009; Lau, Wang, Hsu, & Liu, 2013; Weinstein et al., 2017). Two classical DA drugs for the treatment of PD and schizophrenia, L-dopa and haloperidol, will be introduced in more detail below, as they were also used in the current study. Moreover, DA drugs are widely applied in pharmacological research on animals and humans to investigate the biochemical and functional properties of the DA brain system and its different subcomponents. For example, a wide range of selective DA receptor agonists and antagonists have been developed to target and characterize the different subtypes of the DA receptor and their physiological functions (Missale et al., 1998; Nichols, 2010; Prante et al., 2010; Strange & Neve, 2013). Aside from clinical and research drugs, many recreational drugs exert (part of) their psychotropic properties by targeting specific components of the DA brain system, typically resulting in an enhanced DA transmission (Lüscher & Malenka, 2011; Lüscher & Ungless, 2006; Nestler, 2005; Sulzer, 2011). Such recreational DA drugs include, for example, cocaine (Pomara et al., 2012), amphetamine (Calipari & Ferris, 2013; Fleckenstein, Volz, Riddle, Gibb, & Hanson, 2007), and MDMA, also known as ecstasy (Fleckenstein et al., 2007; Kalant, 2001).

One widely used DA drug for the medical treatment of PD and for research is L-dopa, also known as levodopa. Chemically, L-dopa is the levorotatory isomer of the non-proteinogenic amino acid 3,4-dihydroxyphenylalanine and an immediate precursor of DA (see 1.2.1). In the biosynthetic pathway of catecholamines, L-dopa itself is produced from the amino acid L-tyrosine and then directly converted to DA through decarboxylation by the enzyme DDC (see Figure 2; Meiser et al., 2013).

For treatment and research, synthetically produced L-dopa is used as a drug to increase DA concentrations in the brain by providing additional substrate for DA biosynthesis (Hälbig & Koller, 2007; Pryor & Storer, 2013). In the clinical field, this approach is known as "dopamine replacement therapy" (see below), describing a treatment which compensates for the lack of endogenous DA in PD patients. Dopamine itself is not a suitable drug for such a treatment, as it cannot pass the blood-brain barrier, whereas L-dopa is actively transported into the central nervous system (CNS) via the large neutral amino acid (LNAA) transport system (Contin & Martinelli, 2010; Wang et al., 2017). L-dopa is usually administered in combination with a peripheral DDC inhibitor (DDCI; e.g. benserazide or carbidopa) to avoid its conversion to DA in the peripheral tissue. Thereby, co-administration of a DDCI largely increases the cerebral bioavailability of L-dopa and prevents adverse side effects caused by excessive peripheral DA, such as nausea and hypotension (Hauser, 2009; Nord, 2017; Pryor & Storer, 2013). One frequently used L-dopa/DDCI combination product for PD treatment and research is known under the brand name Madopar (L-dopa/benserazide, 100/25 mg), which has also been used in the current study. Once L-dopa has crossed the blood-brain barrier, it is rapidly metabolized to DA within the CNS by the endogenous enzyme DDC, whereby the exact brain sites of this reaction are unknown (see Hälbig & Koller, 2007). Previous studies in animals and humans indicate that the highest DDC concentrations are located in the terminals of nigrostriatal DA neurons (Hälbig & Koller, 2007; Hefti & Melamed, 1980; Lloyd, Davidson, & Hornykiewicz, 1975; Lloyd & Hornykiewicz, 1972; Melamed, Hefti, & Wurtman, 1980). Still, also other sites exhibit DDC activity and are likely to contribute to the conversion of exogenous L-dopa to DA, including capillary endothelial cells and the terminals of serotonergic and noradrenergic neurons (Bertler, Falck, Owman, & Rosengrenn, 1966; Hefti & Melamed, 1980; Hökfelt, Fuxe, & Goldstein, 1973; Kitahama et al., 2009; Melamed et al., 1980; Mura, Jackson, Manley, Young, & Groves, 1995; Ugrumov, 2009). This newly synthesized, exogenous DA can then contribute to DA neurotransmission within the CNS by binding to pre- and postsynaptic DA receptors in addition to the endogenous DA (Hälbig & Koller, 2007; Pryor & Storer, 2013). More specifically, previous research suggests that exogenous L-dopa is taken up by nigrostriatal nerve terminals, converted to DA, stored in synaptic vesicles, and then co-released with endogenous DA upon neural excitation, resulting in increased striatal DA release (see Breitenstein et al., 2006; Cools, 2006; Floel et al., 2008; Hälbig & Koller, 2007; Horne, Cheng, & Wooten, 1984). Hence, by this mechanism, L-dopa administration may transiently normalize (striatal) DA levels in DA-depleted subjects like PD patients or boost DA levels in healthy subjects. Yet, these acute drug effects are only of short duration, since conventional L-dopa/DDCI preparations have a short plasma half-life of 60 to 90 min and reach peak plasma concentration about 30 to 60 min after oral ingestion (Baruzzi et al., 1987; Iwaki et al., 2015; Keller et al., 2011; Nyholm et al., 2012; see also reviews by Contin & Martinelli, 2010; Hälbig & Koller, 2007; Khor & Hsu, 2007). During chronic L-dopa treatment, however, the duration of the drug response can by far exceed the short half-life of the drug. Especially early stage PD patients often experience improved motor symptoms for several days after L-dopa administration, termed "long-duration response", in addition to a "short-duration response" that parallels L-dopa

plasma concentrations (Anderson & Nutt, 2011; Khor & Hsu, 2007; Nagao et al., 2018; Nutt, Carter, & Woodward, 1995; Zappia & Nicoletti, 2010). Degradation of L-dopa follows principally the same pathways as for endogenous L-dopa (Hälbig & Koller, 2007), but is shifted in the presence of DDCIs towards peripheral COMT metabolism in the liver and kidney, whereby metabolites are predominantly eliminated in the urine (Contin & Martinelli, 2010; Khor & Hsu, 2007; Männistö & Kaakkola, 1999).

Clinically, L-dopa is used for dopamine replacement therapy since the 1960s and still remains the gold standard for the pharmacological treatment of PD (for reviews see Hauser, 2009; Hornykiewicz, 2002, 2017; Ovallath & Sulthana, 2017). It has to be noted, however, that all currently available antiparkinsonian drugs – including L-dopa – only provide symptomatic treatment with no proven protective or curative effect on the disease (see Fahn, Jankovic, & Hallett, 2011). Still, clinical trials have shown L-dopa to be the most effective and best tolerated drug for symptomatic PD therapy (Fahn, 2006; Katzenschlager & Lees, 2002; Müller, 2007; Poewe, Antonini, Zijlmans, Burkhard, & Vingerhoets, 2010). Especially in the early stages of treatment, L-dopa largely alleviates motoric symptoms of PD like bradykinesia, rigidity, and tremor (Bernheimer, Birkmayer, Hornykiewicz, Jellinger, & Seitelberger, 1973; Hälbig & Koller, 2007; Poewe et al., 2010; Pryor & Storer, 2013). On the other hand, L-dopa can also have adverse side effects like nausea, hypotension, and sedation (Hälbig & Koller, 2007; Pryor & Storer, 2013). Furthermore, L-dopa therapy can cause neuropsychiatric symptoms, such as hallucinations, paranoid psychosis, and impulse control disorders (ICDs), the latter including pathological gambling, hypersexuality, and binge eating (Hälbig & Koller, 2007; O'Sullivan, Evans, & Lees, 2009; Weintraub et al., 2010; see also 6.2.1.1). In the long term, chronic L-dopa treatment also leads to serious motoric complications in most PD patients, including involuntary movements (dyskinesia) and “wearing off” phenomena, meaning that medication effects become progressively shorter during treatment, presumably owing to the reduced DA storage capacity with progressing loss of DA neurons (Bhidayasiri & Truong, 2008; Dewey, 2004; Hälbig & Koller, 2007; Jankovic, 2005). Wearing off is followed by stronger and more unpredictable fluctuations in the drug response called “on-off” phenomenon, which makes it difficult to adequately control motoric symptoms in the later stages of treatment (Bhidayasiri & Tarsy, 2012; Dewey, 2004; Hälbig & Koller, 2007; Jankovic, 2005). These severe motoric complications have been shown to emerge in 50 % of PD patients after five years of L-dopa treatment and in > 80 % of patients after 10 years of treatment (see Dodel, Berger, & Oertel, 2001; Hälbig & Koller, 2007) and to strongly impair the patients’ quality of life (Chapuis, Ouchchane, Metz, Gerbaud, & Durif, 2005; Dodel et al., 2001; Gómez-Esteban et al., 2007; Sławek, Derejko, & Lass, 2005). Therefore, current research aims to develop new L-dopa formulations with enhanced pharmacological properties (Contin & Martinelli, 2010; Haddad et al., 2017; Hauser, 2009; Ovallath & Sulthana, 2017), as well as alternative PD treatment strategies, including other drugs like DA agonists, COMT inhibitors, and MAO inhibitors (Fahn et al., 2011; Kaakkola, 2000; Pryor & Storer, 2013; Riederer & Laux, 2011) in addition to neurosurgical approaches (Benabid et al. 2009; Fang & Tolleson, 2017).

Another frequently used DA drug for therapy and research is the DA receptor antagonist haloperidol, which is a common antipsychotic (neuroleptic) medication for the treatment of schizophrenia (see below). Haloperidol is an organic substance belonging to the class of butyrophenones, which are derived from the aromatic compound butyrophenone (1-phenylbutan-1-one). The two halogen atoms in its structure, fluorine and chlorine, have led to its generic name haloperidol (Janssen et al., 1959; López-Muñoz & Alamo, 2009). Although the drug's therapeutic mechanism is not completely understood, it appears to primarily rely on the competitive blockade of DA receptors in the mesolimbic DA system (Brenner & Stevens, 2013; Kapur, Agid, Mizrahi, & Li, 2006; Labbate, 2010; Pryor & Storer, 2013), meaning that DA receptors are occupied but not activated by the drug. Haloperidol thereby exhibits the highest affinity for the D2 receptor (dissociation constant $K_d = 1$ nM) and other D2-like receptors like D3 ($K_d = 5$ nM) and D4 ($K_d = 2$ nM), but lower affinity for the D1 receptor ($K_d = 25$ nM) and D5 receptor ($K_d = 12$ nM; Bymaster et al., 1999). Notably, haloperidol not only blocks postsynaptic but also presynaptic D2 autoreceptors, which can result in an increase rather than decrease of DA transmission, specifically at acute low doses (see 6.2.2.1; Frank & O'Reilly, 2006; Starke, Göthert, & Kilbinger, 1989; Westerink, 2002). Moreover, haloperidol also antagonizes other neurotransmitter receptors like the adrenergic α1 receptor ($K_d = 46$ nM) and the serotonergic 5-HT_{2A} receptor ($K_d = 58$ nM), contributing to the drug's side effect profile (Bymaster et al., 1999; Pryor & Storer, 2013). However, compared to other typical antipsychotics, haloperidol only shows negligible affinity ($K_d > 1000$ nM) for the histamine H1 receptor and muscarinic acetylcholine receptors, thereby exhibiting less antihistaminic and anticholinergic side effects like sedation or weight gain (Bymaster et al., 1999; Labbate, 2010; Li, Snyder, & Vanover, 2016). Haloperidol can be administered orally, intravenously, or intramuscularly to human subjects (Kudo & Ishizaki, 1999). As a highly lipophilic substance, it can freely distribute into different tissues and can also cross the blood-brain-barrier to enter the brain (D'Ambrosio, Zivkovic, & Bartholini, 1982; Kudo & Ishizaki, 1999; Labbate, 2010; Schinkel, Wagenaar, Mol, & van Deemter, 1996). The pharmacokinetic parameters of orally administered haloperidol show a high variability between different studies and subjects: Reported mean values for the plasma half-life range between 14.5 to 36.7 hours, and for the time of peak plasma concentration between 1.7 and 6.1 hours (see review by Kudo & Ishizaki, 1999). Haloperidol is extensively metabolized in the human liver by cytochrome P450 enzymes to various substances, which are primarily eliminated from the body in the urine and bile (Kudo & Ishizaki, 1999; Li et al., 2016; Pryor & Storer, 2013).

In the clinic, haloperidol is used as an antipsychotic drug since the late 1950s and is primarily marketed under the trade name Haldol (Li et al., 2016; López-Muñoz & Alamo, 2009). It is a prototypical example for the class of first-generation (or "typical") antipsychotics and one of the most frequently used drugs in the treatment of schizophrenia and other psychotic disorders (Dold, Samara, Li, Tardy, & Leucht, 2015; Kudo & Ishizaki, 1999; Li et al., 2016). Haloperidol is especially effective for the treatment of positive symptoms like hallucinations and delusions, but is also used for relapse prevention and long-

term stabilization (Geddes, 2002; Labbate, 2010). In contrast, negative symptoms during the chronic phase of schizophrenia, as well as cognitive and executive dysfunctions respond less well to the treatment with haloperidol or other typical antipsychotics (Labbate, 2010; Li et al., 2016). The full therapeutic effects of the drug occur only after weeks of treatment, suggesting that these long-term effects reflect secondary or adaptive responses to D2 antagonism, possibly relying on altered gene expression and synaptic reorganization (Brenner & Stevens, 2013; Labbate, 2010). On the other hand, haloperidol also produces serious adverse side effects on various organ systems, the most characteristic being diverse movement disorders known as "extrapyramidal symptoms" (EPS; see Haddad, Das, Keyhani, & Chaudhry, 2012; Haddad & Dursun, 2008; Leucht et al., 2013; Pryor & Storer, 2013). EPS include dystonia (involuntary muscle contractions), akathisia (motor restlessness), and PD-like symptoms like bradykinesia, rigidity, and tremor. These motoric side effects appear to be primarily caused by D2 receptor blockade in the nigrostriatal pathway, although the precise mechanism is unclear (Kapur, Zipursky, Jones, Remington, & Houle, 2000; Labbate, 2010; Pryor & Storer, 2013). Furthermore, haloperidol can also cause potentially lethal side effects, such as the neuroleptic malignant syndrome (NMS; see Berman, 2011; Haddad & Dursun, 2008) and cardiac adverse reactions like arrhythmias and sudden cardiac death (Girardin & Sztajzel, 2007; Leonard et al., 2013; Ray, Chung, Murray, Hall, & Stein, 2009). Because of the drug's various side effects, its narrow therapeutic window, and the wide interindividual variation in its pharmacokinetics, haloperidol treatment needs to be individually optimized and carefully monitored (Kapur et al., 2000; Kudo & Ishizaki, 1999). An alternative to haloperidol for the treatment of schizophrenia are the newer second-generation (or "atypical") antipsychotics, such as clozapine and risperidone. These drugs exhibit a different receptor affinity profile with lower D2 selectivity and are claimed to have higher efficiency against negative symptoms and a reduced risk of EPS (see Labbate, 2010; Leucht et al., 2009; Li, Snyder, & Vanover, 2016). However, recent meta-analyses comparing the therapeutic efficiencies and side effect profiles of several antipsychotics have not convincingly shown a general superiority of the second- over the first-generation drugs and, moreover, challenge the straightforward classification of antipsychotics into these two distinct categories (Davis, Chen, & Glick, 2003; Geddes, Freemantle, Harrison, & Bebbington, 2000; Leucht et al., 2009, 2013; see also Labbate, 2010; Pryor & Storer, 2013).

Both drugs, L-dopa and haloperidol, are often used in human research to investigate the causal effects of increased or decreased DA transmission on different behaviors, including reward-based decision making (e.g. Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Pine, Shiner, Seymour, & Dolan, 2010; Wunderlich, Smittenaar, & Dolan, 2012). Exemplary for such an approach is the pharmacological fMRI experiment of Pine et al. (2010), who used a placebo-controlled, double-blind, within-subjects design to study temporal discounting behavior in healthy subjects under both dopaminergic (L-dopa) and antidopaminergic (haldoperidol) drug conditions. Notably, a similar experimental design has been employed in the current study to test for DA-dependent effects on explore/exploit behavior (see 1.4.2). However, this section has also demonstrated that the mechanisms underlying the seemingly opposite

drug actions of L-dopa and haloperidol are rather complex, with various factors contributing to the (anti-)dopaminergic effects of these drugs. Hence, this mechanistic complexity should be kept in mind when utilizing DA drugs like L-dopa and haloperidol for research.

To summarize, DA drugs provide an important tool for both the treatment of DA-related diseases and for neuroscientific research. In the clinic, the DA precursor L-dopa is a widely used drug for dopamine replacement therapy in PD, while the D2 receptor antagonist haloperidol is a typical antipsychotic drug for the treatment of schizophrenia. For research, L-dopa and haloperidol are often used to examine the behavioral and neural effects of either increased or decreased DA transmission, respectively, an approach that was also employed in the current study.

1.2.4 Inverted-U hypothesis of DA

The "inverted-U hypothesis" of DA states that the relationship between DA levels and cognitive performance follows an inverted-U-shaped function, or optimum curve, in which deviations from optimal DA levels in both directions result in a deterioration of performance (see Figure 3; Cools & D'Esposito, 2011). In other words, both too low (depleted) and too high (excessive) DA levels impair cognitive functioning, whereas moderate DA levels allow for an optimal performance. Several studies on animals and humans have, in general, provided empirical support for this hypothesis and its implications (reviewed by Cools & D'Esposito, 2011; Floresco, 2013).

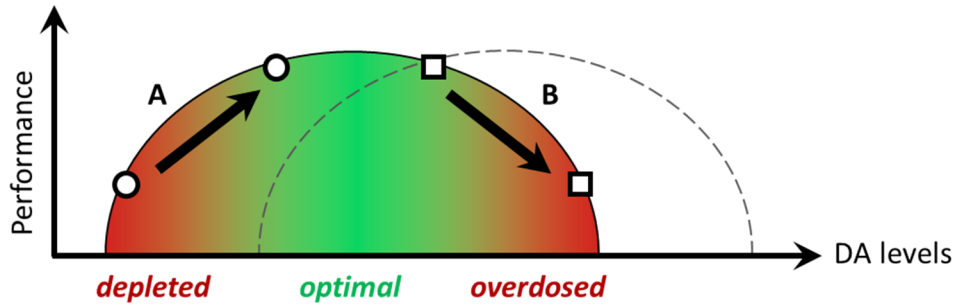


Figure 3. Inverted-U-shaped function of dopamine (DA). The relationship between DA levels and cognitive performance follows an inverted-U curve, where both too little (depleted) and too much (overdosed) DA impairs performance. Hence, the same DA drug should produce opposite effects on performance in two subjects with different baseline DA levels (A and B). Note also that optimal DA levels might vary between different tasks and brain regions, as indicated by the shifted (dashed) curve. See text for further explanations. Adapted from Cools & D'Esposito (2011).

A first implication of the inverted-U hypothesis is that subjects with very low or very high baseline DA levels perform worse in DA-related cognitive tasks than subjects with intermediate DA levels. Note that the term "baseline DA level" thereby refers to the DA level of the subject without drug administration or other external DA manipulations. Previous empirical studies have confirmed this relationship for different cognitive functions (Akbari Chermahini & Hommel, 2010, 2012; Dang, Xiao, Liu, Jiang, & Mao, 2016; Ueda, Tominaga, Kajimura, & Nomura, 2016; see also review by Jongkees &

Colzato, 2016). For example, Akbari Chermahini & Hommel (2010) used a divergent thinking task to assess creative performance in humans, which is claimed to rely on DA. In a large sample of 117 subjects, they showed that the baseline DA level – as indexed by the spontaneous eye blink rate (sEBR; see 1.2.5) – predicts individual task performance according to an inverted-U-shaped function, with medium sEBRs predicting optimal performance. Later studies have reported similar inverted-U-shaped relationships between the sEBR and creative thinking (Akbari Chermahini & Hommel, 2012; Ueda et al., 2016) or self-regulatory control (Dang et al., 2016). In addition, the idea that non-optimal DA levels are associated with cognitive impairment also contributed to a better understanding of the cognitive symptoms observed in diseases linked to excessive or depleted DA levels, such as Parkinson's disease (Williams-Gray, Hampshire, Barker, & Owen, 2008; Wu et al., 2012), schizophrenia (Alawieh et al., 2012; Kömek, Bard Ermentrout, Walker, & Cho, 2012; Wu et al., 2012), attention deficit hyperactivity disorder (ADHD; Levy, 2009), and obsessive-compulsive spectrum disorders (van Velzen, Vriend, de Wit, & van den Heuvel, 2014). Also, cognitive impairments in healthy subjects due to aging or acute stress have been linked to depleted or excessive DA levels, respectively, based on the inverted-U hypothesis of DA (Goldman-Rakic, Muly, & Williams, 2000; Seamans & Yang, 2004; Williams & Castner, 2006).

A second implication of the inverted-U hypothesis is that administration of a DA drug can have mixed and even opposite effects in different subjects, depending on their individual baseline DA level. As demonstrated in Figure 3, a drug-induced increase in DA levels (e.g. by L-dopa) should improve cognitive performance in individuals with low DA levels, but impair performance in subjects with already optimal or above-optimal DA levels. In contrast, a drug-induced decrease in DA levels (e.g. by haloperidol) should improve cognitive functioning in subjects with high DA levels, but have detrimental effects in individuals with optimal or below-optimal DA levels. Thus, administration of the same drug can lead to opposite (paradoxical) drug effects in subjects with low compared to high baseline DA levels. Several studies in animals and humans have provided empirical evidence for such paradoxical drug effects, thereby supporting the inverted-U hypothesis of DA (see reviews by Cools & D'Esposito, 2011; Floresco, 2013; Jongkees & Colzato, 2016; Schacht, 2016). For instance, Cools et al. (2009) showed that the D2 receptor agonist bromocriptine improved reward-based (relative to punished-based) reversal learning in human subjects with low striatal DA synthesis capacity, but impaired it in subjects with high striatal DA synthesis capacity, as measured by positron emission tomography (PET). Note that differences in striatal DA synthesis capacity were suggested to reflect differential baseline levels of synaptic DA in this study. Other studies found a similar DA baseline dependency of DA drug effects on other cognitive functions, including working memory (Frank & O'Reilly, 2006; Gibbs & D'Esposito, 2005, 2006; Mehta et al., 2000), set shifting (Frank & O'Reilly, 2006; Kimberg, D'Esposito, & Farah, 1997), and response conflict processing (Cavanagh, Masters, Bath, & Frank, 2014). Instead of PET imaging, however, these studies assessed individual baseline DA levels by the use of behavioral proxies like the sEBR and working memory capacity (see 1.2.5). Moreover, additional evidence stems

from human pharmacogenetic studies showing that DA drug effects are strongly modulated by certain genotypes that are indicative of a subject's baseline DA level, such as the COMT Met/Val polymorphism (reviewed by Schacht, 2016). For example, one such study (Farrell, Tunbridge, Braeutigam, & Harrison, 2012) showed that an increase in prefrontal DA levels via the COMT inhibitor tolcapone increases working memory performance and risk aversion in Val/Val subjects (with supposedly low baseline DA levels in the PFC), but has the opposite effect in Met/Met subjects (with supposedly higher baseline DA levels in the PFC). The review by Schacht (2016) suggests, however, that such pharmacogenetic effects depend on the subcomponent of the DA system that is targeted by the drug: While there is strong evidence for pharmacogenetic effects with antipsychotic drugs, which mainly target D2 receptors, the evidence is mixed for psychostimulants and COMT inhibitors, which exhibit larger effects on D1 receptor transmission. Finally, it was shown that the individual baseline DA level does not only modulate DA drug effects, but also the effects of tDCS neurostimulation on cognitive performance (reviewed by Wiegand, Nieratschker, & Plewnia, 2016). For instance, two human tDCS studies found that anodal (excitatory) stimulation of the PFC reduced executive functioning specifically in Met/Met subjects (Plewnia et al., 2013), while cathodal (inhibitory) PFC stimulation reduced executive functioning specifically in Val/Val subjects (Nieratschker, Kiefer, Giel, Krüger, & Plewnia, 2015). Both results are in line with the inverted-U hypothesis, assuming that Val homozygotes are located on the ascending (left) side of the curve where DA inhibition decreases performance, while Met homozygotes are located on the descending (right) side of the curve where DA stimulation decreases performance (see Wiegand et al., 2016).

Although the inverted-U hypothesis of DA has gained much empirical support, research has also shown that the true relationships are probably much more complex than the simple idea suggests, with many important questions remaining open. First of all, the observation of an inverted-U-shaped relationship between DA levels and cognitive performance is, at present, only a "descriptive rather than a mechanistic account of the action of DA" (Cools & D'Esposito, 2011, p. e121). It remains largely unknown, by which neural mechanisms too low and too high DA levels impair cognitive functioning and if (or how) these two extremes differ (for mechanistic speculations see Cools & D'Esposito, 2011; Floresco, 2013). Furthermore, the inverted-U hypothesis remains relatively vague about the exact meaning of the term "baseline DA levels". Since DA concentrations cannot be directly measured in the living human brain, most research on the inverted-U hypothesis relies on behavioral or genetic proxies to assess baseline DA levels, whereby it is often unclear which specific aspect of DA function these proxies reflect (see 1.2.5). For instance, it remains unsettled whether the inverted-U function rather applies to the modulating effects of striatal DA (Cools et al., 2009) and/or prefrontal DA (Floresco, 2013; Floresco & Magyar, 2006; Mattay et al., 2003; Schacht, 2016), or perhaps to frontostriatal connectivity (Cools & D'Esposito, 2011). Pharmacological research on animals suggests, however, that the inverted-U function specifically describes the relationship between prefrontal D1 receptor activity and working memory performance, whereas the relations between other DA receptor subtypes and

cognitive domains may follow different functions (Floresco, 2013; Floresco & Magyar, 2006; see also Fallon et al., 2015). Based on this evidence, Floresco (2013) concludes that the inverted-U curve is not a "one-size-fits-all function", but rather represents one specific member in a family of functions describing how DA modulates behavior across different cognitive domains. Finally, it was suggested that the exact shape of the inverted-U curve may be highly variable between subjects and tasks (see Wiegand et al., 2016), and also that the position of the curve's turning point may vary between different tasks and brain regions to exhibit distinct task- and region-specific DA optima (Cools & D'Esposito, 2011; Fallon et al., 2015). Regarding these points, more research is clearly needed to specify when and in what form the inverted-U function applies to different cognitive domains, as well as to reveal the neural mechanisms underlying that relationship.

To conclude, while many aspects of the inverted-U hypothesis remain open and require more research, the hypothesis has proven highly useful to describe and predict how individual baseline DA levels modulate DA-dependent cognitive functions and DA drug effects. Consequently, further research on DA-related behaviors should take individual differences in baseline DA levels into account and also directly test for non-linear (quadratic) relationships as predicted by the inverted-U hypothesis.

1.2.5 DA proxies

Currently, there are no techniques to directly measure intra- or extracellular DA concentrations in the living human brain (see Badgaiyan, 2014). While certain aspects of DA function – such as DA synthesis capacity and DA receptor availability or occupancy – can be visualized *in vivo* using PET imaging with radiolabeled receptor ligands or metabolites (Badgaiyan, 2011; Laruelle, 2000), this technique is expensive and methodically complex, partly limiting its use for research. Therefore, most DA research relies on the use of behavioral or genetic proxy measures to indirectly assess central DA function, providing a cheap and non-invasive alternative to PET. In the following, two of these proxy measures, the spontaneous eye blink rate and the working memory capacity, are introduced in more detail, as they were both used in the current study.

The spontaneous eye blink rate (sEBR) is one of the oldest and most widely used proxy measures for central DA function. It is usually measured under resting conditions ("tonic sEBR") by counting the number of spontaneous eye blinks over a course of several minutes, e.g. through direct observation, video recording, or electromyography (see 2.3.1). Note, however, that the sEBR can also be measured in response to stimulus conditions, such as a cognitive task or a video, and is then referred to as "phasic sEBR". The utility of the sEBR as an indicator of DA function has recently been evaluated in an extensive research review by Jongkees and Colzato (2016). Based on more than 100 studies, the review concluded that the sEBR is a "useful predictor of DA in a wide variety of contexts" (p. 58), whereby three main research contexts have been distinguished. First, several pharmacological studies have investigated the effects of different DA drugs on the sEBR in animals and humans. Overall, these studies showed that the sEBR can reflect drug-induced changes in both D1 and D2 receptor activity, with higher

sEBR indicating higher DA function. However, when measured under non-pharmacological (baseline) conditions, the sEBR was found to be positively related to D2 but not D1 receptor availability (Groman et al., 2014). Hence, it has been suggested that the sEBR may reflect D1 receptor activity only under pharmacological conditions, which might be explained by the lower DA sensitivity of D1 compared to D2 receptors (see Jongkees & Colzato, 2016). Second, the sEBR has been widely used to study DA dysfunctions in various human disorders, both at baseline and after DA drug treatment. For example, it was found that the sEBR is typically reduced in PD patients and increased in schizophrenic patients, consistent with the notion that these patients exhibit diminished or excessive striatal DA function, respectively (e.g. Adamson, 1995; Aksoy, Ortak, Kurt, Cevik, & Cevik, 2014; Bologna et al., 2012; Deuschl & Goddemeier, 1998; Karson, 1983; Mackert, Woyth, Flechtner, & Frick, 1988; Stevens, 1978; see also review by Jongkees & Colzato, 2016). Moreover, while DA-stimulating treatment (e.g. with L-dopa) typically increases the sEBR in PD patients (Agostino et al., 2008; Bologna et al., 2012; Kimber & Thompson, 2000), neuroleptic treatment reduces the sEBR in schizophrenic patients (Adamson, 1995; Bartkó, Herczeg, & Zádor, 1990; Karson, 1983; Karson et al., 1981; Kleinman et al., 1984), both effects indicating a (partial) normalization of DA function in response to the DA drugs. Third, many human studies have used the sEBR to examine the relationship between individual differences in baseline DA function and cognitive performance. Overall, these studies have shown that the sEBR at baseline can reliably predict individual differences in task performance, particularly in tasks of cognitive flexibility and reinforcement learning (see Jongkees & Colzato, 2016). While some of these studies found a linear relation between sEBR and performance (e.g. Slagter, Georgopoulou, & Frank, 2015; Zhang et al., 2015), other studies reported an inverted-U-shaped relationship (see 1.2.4; e.g. Akbari Chermahini & Hommel, 2010, 2012; Ueda et al., 2016). Although the sEBR has been extensively used across all these research contexts, it is presently unknown by what mechanism the relationship between sEBR and central DA function can be explained. One hypothesis is that DA modulates – via the basal ganglia – the activity of the spinal trigeminal complex, which has been suggested to be an integral part of the spontaneous blink generator circuit (Kaminer, Powers, Horn, Hui, & Evinger, 2011; Kaminer, Thakur, & Evinger, 2015). Finally, some limitations of the sEBR include its (relative) unspecificity for different DA subfunctions, such as different DA pathways and DA receptor systems, and the large methodological variability in assessing the sEBR (see Jongkees & Colzato, 2016). In addition, recent PET studies in humans have questioned the validity of the sEBR as a (positive) predictor of DA, as they have found either no or even a negative relationship between the sEBR and different aspects of central DA function (Dang et al., 2017; Sescousse et al., 2018).

A second behavioral marker for baseline DA function is the working memory capacity (WMC), which has gained attention as a DA proxy measure over the last two decades (see Cools & D'Esposito, 2011). It is usually measured with working memory span tasks, such as the reading span test (Daneman & Carpenter, 1980) or the listening span test (Daneman & Carpenter, 1980; Salthouse & Babcock, 1991). In these two tests, subjects read aloud or listen to a series of sentences and need to recall the final

word of each sentence in the correct order. Notably, these and other WMC tasks are typically dual-task paradigms, which require both the storage of information (e.g. memorizing words) and the simultaneous processing of information (e.g. reading or listening), in contrast to simple short-term memory tasks which only require the storage of information (see Conway et al., 2005; Kail & Hall, 2001; Kane et al., 2004; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Unsworth & Engle, 2007). Both the reading and listening span test have been used in several studies as a proxy for baseline DA function (e.g. Cools et al., 2009; Cools, Gibbs, Miyakawa, Jagust, & D'Esposito, 2008; Gibbs & D'Esposito, 2005, 2006; Kimberg et al., 1997; Kimberg & D'Esposito, 2003; Landau, Lal, O'Neil, Baker, & Jagust, 2009). Additionally, other studies have also used the digit span task to measure baseline WMC as a proxy for DA (Wechsler, 2008; e.g. used by Mehta et al., 2000; van der Schaaf et al., 2014). In this task, a series of numerical digits is presented to the participant and needs to be recalled in the same order (forward version) or in the reverse order (backward version). Note that while the forward version is considered a measure of short-term memory rather than WMC (Kail & Hall, 2001; Kane et al., 2004; Unsworth & Engle, 2007), the backward version requires both the simultaneous storage and processing of information (i.e. recalling and reordering the digit string) and has hence been regarded and used as a measure of WMC (see Conway et al., 2005; Mehta et al., 2000; Oberauer et al., 2000). The first direct evidence for an association between WMC and baseline DA function stems from a human PET study by Cools et al. (2008). In a small sample of 11 female subjects, this study showed that the WMC, as measured with the listening span test, positively predicts DA synthesis capacity in the striatum. This finding was later replicated in a sample of 22 healthy elderly subjects by another human PET study (Landau et al., 2009). Note, however, that both studies did not allow to test for an association of WMC with prefrontal DA, as their applied PET technique, 6-[^{18}F]fluoro-L-m-tyrosine (FMT) PET, is optimized to visualize DA function in the striatum, but not in the PFC (see Cools & D'Esposito, 2011; Jordan et al., 1997). Aside from these PET studies, further (indirect) evidence for an association between WMC and baseline DA stems from pharmacological studies showing that baseline WMC predicts DA drug effects on cognitive performance according to an inverted-U-shaped function (see 1.2.4). Several studies have shown, for example, that DA receptor agonists improve cognitive performance in subjects with low WMC, but impair performance in subjects with high WMC (reviewed by Cools & D'Esposito, 2011). This WMC baseline effect was found for different cognitive functions, including set shifting (Frank & O'Reilly, 2006; Kimberg et al., 1997), working memory updating (Frank & O'Reilly, 2006; Mehta et al., 2000), and working memory retrieval (Gibbs & D'Esposito, 2005, 2006). These findings suggest that the individual WMC (low or high) reflects differential baseline levels of DA (low or high), which in turn modulate DA drug effects in opposing ways as predicted by the inverted-U hypothesis of DA (see 1.2.4). One limitation of the WMC is that it is presently not clear which aspect of DA function the proxy reflects, e.g. whether it specifically indicates striatal DA or also prefrontal DA, and to which DA receptor subtype(s) it corresponds. Also, there are currently only a few (small-sample) studies in the literature showing direct associations between WMC and baseline DA, wherefore more research is needed to establish the WMC as a valid and reliable marker of DA function.

1.3 Dopamine in the explore/exploit trade-off

While the preceding sections have broadly introduced the DA brain system and outlined its basic functions in reward-based decision making, this section will now specifically focus on the role of DA in regulating the explore/exploit trade-off based on evidence from animals, humans, and neural network models.

1.3.1 Animal and neural network studies

One line of evidence for a dopaminergic involvement in the explore/exploit trade-off stems from research on animals and neural network models. A main contribution in this field has come from the experimental and theoretical work by Beeler and colleagues. In 2010, Beeler et al. conducted an experiment on dopamine transporter (DAT) knockdown mice, which are characterized by increased levels of tonic DA due to a reduced reuptake of extracellular DA (Zhuang et al., 2001). Compared to wild-type controls, these hyperdopaminergic mice showed less exploitative (i.e. more exploratory) behavior in an instrumental learning task, in which food was earned by lever presses. More specifically, the mice could choose between two levers, one cheap and one expensive lever yielding food rewards at different costs (i.e. number of presses), whereby the assignment of the two levers as cheap or expensive switched from time to time. Behavioral analysis showed that the hyperdopaminergic mice choose the expensive lever more often than control mice, despite responding to switches in lever costs similarly quickly. These findings suggest that DAT knockdown mice are not impaired in learning about changes in the environment, but exert more effort than controls for earning a given amount of reward. In other words, the hyperdopaminergic mice show a lower tendency to base their choices on previously learned reward (cost) rates, i.e. a lower tendency to exploit. Furthermore, a model-based analysis of choice behavior revealed a selective decrease in the softmax β parameter, but not in the learning rate, for the DAT knockdown mice compared to controls, reflecting a noisier, more exploratory choice behavior. Based on the results of this and other studies, Beeler et al. (2012) developed a new conceptual framework of DA function, suggesting that the primary role of DA in behavior is the regulation of energy expenditure or thrift (see also Beeler, 2012). According to this framework, DA modulates energy expenditure along two dimensions: a conserve-expend axis and an explore-exploit axis. The conserve-expend axis reflects the regulation of the general activity level, i.e. how much energy to expend, ranging from low activity (conserve) to high activity (expend). The explore-exploit axis describes instead how to allocate this energy to different behavioral activities along a continuum from exploitation to exploration. Along these two dimensions, behavior is modulated by DA according to the environmental energy conditions: In rich environments with plentiful resources, increased DA drives behavior towards energy expenditure and exploration, whereas in poor environments with scarce resources, decreased DA promotes energy conservation and exploitation. In fact, this framework suggests that the well-established role of DA in reward-related and motivated behaviors may actually arise as a consequence of DA's primary role in behavioral energy management.

Although influential, the theoretical framework of Beeler et al. (2012) remained relatively vague about the exact neural mechanisms underlying the regulation of energy expenditure and explore/exploit behavior by DA. As Beeler et al. (2012) remarked, the term "DA function" in their skeletal framework is broadly construed and may include several factors like extracellular DA levels, tonic and phasic DA cell firing, properties of DA synthesis and vesicular DA release, and the relative expression rates of different DA receptors and DA transporters. A more mechanistic account on the role of DA in explore/exploit regulation has been provided by studies based on neural network models (Humphries et al., 2012; Mandali, Rengaswamy, Chakravarthy, & Moustafa, 2015). For instance, Humphries et al. (2012) performed a simulation experiment on a computational model of the full basal ganglia (BG) circuit to test the hypothesis that tonic striatal DA controls the explore/exploit trade-off via the basal ganglia. The outcome of their simulations supported this hypothesis, showing that variations in striatal DA modulate the BG output in a way favoring either exploratory or exploitative actions. More specifically, they showed that higher tonic DA levels in the striatum lead to a more peaked probability distribution for action selection (i.e. a higher softmax β) as encoded in the BG output, reflecting a choice behavior that is more value-driven (exploitative) and less noisy (exploratory). The authors concluded from these results that tonic DA variations in the striatum are sufficient to control the explore/exploit trade-off by encoding the degree to which action selection in the BG is influenced by previous rewards. In addition, another simulation experiment on a spiking BG network model also supported the view that DA levels in the BG might regulate explore/exploit behavior, albeit suggesting a different mechanism (Mandali et al., 2015). In short, this study showed that the DA-dependent level of neural synchrony within the BG (i.e. between the subthalamic nucleus and globus pallidus externus) modulates to the level of exploration, as demonstrated on simulated choice behavior in the multi-armed bandit task. Specifically, intermediate DA levels in this model were associated with high neural synchrony and more exploratory behavior, whereas high DA levels led to neural desynchronization and more exploitative behavior. Hence, both neural network studies arrive at the conclusion that the BG might represent the subcortical neural substrate for controlling the trade-off between exploitation and random exploration, but also indicate that the actual function linking striatal DA levels to behavior might be rather complex.

Aside from these studies on subcortical DA systems, there is also evidence for an involvement of prefrontal DA in different aspects of explore/exploit behavior, including risk preference, working memory, and behavioral flexibility (reviewed by Cools & D'Esposito, 2011; Floresco, 2013; Floresco & Magyar, 2006). For instance, it was shown that pharmacological manipulation of prefrontal D1 and D2 receptor activity affects risky choice behavior in a probability discounting task, in which rats choose between a small certain and a large but uncertain reward (St Onge et al., 2011). More specifically, the study showed that while D1 blockade in the medial PFC decreases risky choices, D1 stimulation and D2 blockade increases risky choices. The authors concluded from these results that prefrontal D1 and D2 receptors play complementary roles in risky decision making. Further, they suggested that balancing

D1/D2 receptor activity might reflect the mechanism by which prefrontal DA assists the trade-off between exploitation of a certain reward and exploration of an uncertain but potentially larger reward. In addition, prefrontal DA receptors have been implicated in working memory and behavioral flexibility (Floresco, 2013; Floresco & Magyar, 2006), functions which have also been related to explore/exploit behavior (Addicott et al., 2017; Beeler et al., 2014). For example, it was found that D1 receptor activity modulates working memory according to an inverted-U-shaped function, with too low or too high activity leading to impaired performance (see Floresco, 2013). As suggested by Addicott et al. (2017), this DA-dependent modulation of working memory could influence the learning rate component of explore/exploit behavior, i.e. the degree to which learned values (and the choices based thereon) are influenced by the previous reward history. Moreover, it was shown that an increased D1 and D2 receptor activity enhances cognitive flexibility and adaptive behavior, potentially by signaling changes in reward contingencies triggered by unexpected outcomes (Floresco, 2013). This DA-dependent cognitive and behavioral flexibility might also represent an essential component of explore/exploit behavior that facilitates adaptive switching between both decision strategies (Beeler et al., 2014). In sum, previous research on animals and neural network models has yielded evidence for an involvement of both striatal and prefrontal DA in the regulation of explore/exploit behavior and related cognitive functions.

1.3.2 Human studies

The role of DA in explore/exploit behavior has also been investigated in humans by a number of behavioral and few neuroimaging studies. A first line of evidence stems from genetic variation studies, which investigated the influence of different DA gene polymorphisms on explore/exploit behavior (Blanco et al., 2015; Frank et al., 2009; Kayser et al., 2015). For example, Frank et al. (2009) found that genetic variations in prefrontal and striatal DA genes predict individual differences in exploration and exploitation, respectively, as measured by the clock task (see 1.1.2). More specifically, they showed that exploitative behavior was influenced by two genes controlling striatal DA function, namely the DRD2 gene predictive of striatal D2 receptor availability (Hirvonen et al., 2004) and the DARPP-32 gene involved in striatal D1 receptor-mediated synaptic plasticity and reward learning (Calabresi et al., 2000; Stipanovich et al., 2008). In contrast, uncertainty-driven exploration was associated with the COMT gene, which is mainly involved in controlling prefrontal DA function (see 1.2.2; e.g. Bilder et al., 2004; Meyer-Lindenberg et al., 2005). Furthermore, model-based analysis of choice behavior revealed a gene-dose effect between the COMT genotype and exploratory behavior. Note, therefore, that the human COMT gene exists in two allelic variants, the “Val” allele and the “Met” allele, whereby the Met allele is associated with lower enzymatic activity and therefore higher prefrontal DA levels (see Bilder et al., 2004). Based on these allelic variants, Frank et al. (2009) showed that the uncertainty-driven exploration parameter in their cognitive model was highest in Met/Met carriers, intermediate in Val/Met carriers, and smallest in Val/Val carriers. This gene-dose effect suggests that increased prefrontal DA function promotes uncertainty-driven exploration in humans, which is consistent with

animal findings showing that increased prefrontal DA receptor activity promotes behavioral flexibility (see above; Floresco, 2013). While this gene-dose effect between COMT genotype and exploration could not be replicated in two later studies (Blanco et al., 2015; Kayser et al., 2015), one of these studies showed that the COMT inhibitor tolcapone, expected to increase prefrontal DA levels, increases uncertainty-driven exploration, although only in Met/Met subjects (Kayser et al., 2015). In contrast, exploitative behavior was not affected by tolcapone, irrespective of the COMT genotype. While the observed drug-genotype interactions are not easy to explain in mechanistic terms (see Kayser et al., 2015), these results nonetheless support the view that prefrontal DA is involved in regulating uncertainty-driven exploration, and demonstrate again that DA drug effects depend on the individual baseline DA level (see 1.2.4). A third human genetic study (Blanco et al., 2015) used a variant of the two-armed bandit task to investigate the influence of the COMT genotype on explore/exploit behavior under different task load conditions. The authors hypothesized that if strategic exploration depends on prefrontal DA, then Met carriers should outperform Val/Val subjects in the explore/exploit task, especially under high task load conditions that strain the limited cognitive resources of the prefrontal system. The behavioral results supported this hypothesis, showing that Met/Met and Met/Val subjects performed better, i.e. selected the best option more often, than Val/Val homozygotes, but only in the high task load condition. Moreover, model-based analysis of choice behavior suggested that Met carriers and Val/Val subjects followed different exploration strategies under high task load conditions: When comparing a naïve choice model capturing only random exploration (softmax rule) against a more complex choice model capturing also uncertainty-driven exploration (Ideal Actor model), Met carriers were more likely to be best characterized by the more complex model than Val/Val subjects. In sum, the reported human genetic studies largely support the notion that prefrontal DA is involved in exploratory behavior, especially in uncertainty-driven exploration.

A second line of (indirect) evidence comes from studies examining the role of DA in risky decision making and cognitive flexibility, which are both believed to represent essential components of explore/exploit behavior (see Addicott et al., 2017; Beeler et al., 2014). Most of these studies also used genetic variations of different DA genes as a proxy for DA function. For example, a number of human studies have found that the COMT Met allele (linked to higher prefrontal DA levels) is associated with better set shifting performance as measured with the Wisconsin Card Sorting Test (WCST; Caldú et al., 2007; Egan et al., 2001; Malhotra et al., 2002; Minzenberg et al., 2006; Rosa et al., 2004). Hence, consistent with the animal literature reported above (see Floresco, 2013), these findings suggest that prefrontal DA promotes cognitive flexibility, which may be related to its role in promoting exploratory behavior (see Beeler et al., 2014). However, it should be noted that this COMT effect on set shifting performance could not be validated in a later meta-analysis including 16 independent studies (Barnett, Scoriels, & Munafò, 2008), questioning the reliability of the earlier findings. Aside from cognitive flexibility, other studies have investigated the influence of DA gene variations on risky decision making

and its neural correlates (Kohno et al., 2016; Lancaster et al., 2012). Risky decision making in these studies was assessed with the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), in which subjects pump a virtual balloon to yield increasingly larger rewards and have to decide in each step to either continue pumping with the risk of losing the reward if the balloon explodes, or to stop the trial with the reward earned so far. Note that this task shows clear similarities to explore/exploit paradigms like the restless bandit task, in which subjects also have to decide in each trial to either safely exploit a known option or to explore a risky option to seek an even larger reward. Using this task, it was shown that COMT Met carriers are more willing to take calculated risks when rewards are attainable than Val/Met or Val/Val subjects (Lancaster et al., 2012), in line with the view that higher prefrontal DA function is associated with increased uncertainty-driven exploration. A second study used a gene composite score as a positive proxy for striatal DA function, which combined functional variations across five different DA genes (Kohno et al., 2016). The study found that this gene composite score predicted BART performance, as measured by the total monetary payout, according to an inverted-U function, and negatively predicted activity in the dorsolateral prefrontal cortex (DLPFC) during risky choices. Hence, these findings suggest that risky decision making, similar to uncertainty-driven exploration, depends on prefrontal substrates, which may reciprocally interact with the striatal DA system during risk/reward (or explore/exploit) decision making (Kohno et al., 2016; see also discussion in 6.2.1.2). Aside from genetic variations, research in this field has also relied on the spontaneous eye blink rate (sEBR) as a proxy for central DA function (see 1.2.5). For instance, it was reported that a higher sEBR (indicating higher baseline DA function) is associated with both enhanced cognitive flexibility (Dreisbach et al., 2005; Müller et al., 2007; Tharp & Pickering, 2011) and higher risk seeking (Sherman & Wilson, 2016). Note, however, that these findings do not allow to conclude which specific aspects of DA function are underlying the observed associations, given that the sEBR is a relatively unspecific proxy for DA function (see 1.2.5). While the sEBR is mostly considered to reflect striatal DA function (see Jongkees & Colzato, 2016), both cognitive flexibility and risky decision making have often been linked to prefrontal DA function (see above; e.g. Floresco, 2013; Floresco & Magyar, 2006; St Onge et al., 2011; St Onge & Floresco, 2010), which might as well be reflected by the sEBR. Hence, when interpreting the sEBR findings in terms of prefrontal DA function, they would be consistent with the human genetic studies reported above, which link higher prefrontal DA function to increased uncertainty-driven exploration (Blanco et al., 2015; Frank et al., 2009; Kayser et al., 2015). In sum, several human studies provide evidence for a DA involvement in risky decision making and cognitive flexibility, two functions that are likely to subserve exploratory behavior.

A third line of evidence stems from research on human disorders associated with a dysregulation of the DA brain system, such as PD and schizophrenia. While PD is characterized by depleted striatal DA levels, schizophrenia is commonly regarded to involve excessive striatal but reduced prefrontal DA function, as described above (see 1.2.3). Hence, both these disorders are characterized by either too

low or too high activity in distinct DA subsystems, which can be used to examine the influence of DA on different behaviors.

On the one hand, research on PD patients found that these patients show altered reward-maximizing (exploitative) behavior in the clock task compared to healthy controls, whereas exploratory behavior was not found to be altered in these patients (Moustafa et al., 2008). More specifically, PD patients off medication were selectively impaired in speeding up their responses to maximize rewards (Go learning), while the same patients on medication (i.e. L-dopa and/or DA agonists) were selectively impaired in slowing down their responses to maximize rewards (NoGo learning). Similar DA medication effects in PD patients have also been reported by Frank, Seeberger, and O'Reilly (2004) in two classical reinforcement learning tasks. Furthermore, both studies showed that the observed DA medication effects on Go/NoGo learning could be simulated using a computational model of the basal ganglia network, hence supporting the view that these effects depend on striatal DA function. Another PD patient study examined DA medication effects on risky decision making and its neural correlates (van Eimeren et al., 2009). This study found that PD patients acutely medicated with the DA agonist pramipexole showed increased risk-taking behavior compared to the same patients off medication, which was associated with drug-induced changes in prefrontal activity. More specifically, a model-based analysis revealed that pramipexole led to a desensitization of the lateral OFC towards negative reward prediction errors. The authors concluded from these results that the drug-induced increase in tonic DA activity may prevent pauses (“dips”) in DA transmission in response to losses, thereby impairing negative reinforcement learning and promoting risk-taking behavior. In contrast, no such effect was observed in the same PD patients acutely medicated with L-dopa, which is considered to stimulate phasic rather than tonic DA transmission (van Eimeren et al., 2009; see also discussion in 6.2.1.1). Taken together, the reported DA medication effects in PD patients are in line with the above findings from human genetic studies (e.g. Frank et al., 2009), supporting the view that striatal DA drives reinforcement learning and exploitation, while prefrontal DA promotes risk taking and uncertainty-driven exploration.

On the other hand, research on explore/exploit behavior in schizophrenic patients found that these patients show impaired positive reinforcement learning (Go learning) and substantially reduced uncertainty-driven exploration compared to healthy controls (Strauss et al., 2011). The authors attributed the deficits in Go learning to a potential dysregulation of subcortical DA in the direct (D1-driven) pathway, which was supported by simulations on a neural network model of the basal ganglia. In contrast, the deficits in uncertainty-driven exploration were attributed to reduced prefrontal DA function. This interpretation was supported by the finding that the effect on directed exploration correlated with the severity of anhedonia, a negative symptom that has been linked to degraded prefrontal DA function (see Abi-Dargham & Moore, 2003; Brisch et al., 2014; Davis et al., 1991; Strauss et al., 2011). Following this interpretation, the reduced exploration in schizophrenic patients might further be regarded as an effect opposite to the above finding in pramipexole-

medicated PD patients, showing that increased prefrontal DA activity promotes risk-taking behavior (van Eimeren et al., 2009). Yet, a concrete mechanistic explanation of these results needs further research, especially since all schizophrenic patients in the sample were already treated with antipsychotic drugs, which might have affected central DA function in these patients in unknown ways. In addition, another study in schizophrenic patients examined instead the role of prefrontal DA in cognitive flexibility (Egan et al., 2001). The study found that cognitive flexibility was significantly reduced in schizophrenic patients compared to healthy controls, whereby healthy siblings of these patients showed intermediate performance between both groups. Moreover, COMT genotyping revealed a higher frequency of the Val allele in schizophrenic patients compared to healthy controls, and again intermediate levels in the patients' siblings. Based on these findings, the authors suggested that the COMT Val allele and its effects on prefrontal DA function may be responsible for the impairments in cognitive function and the increased risk for schizophrenia (see also Weinberger et al., 2001). Hence, the observation that cognitive flexibility is reduced in both schizophrenic patients (Egan et al., 2001) and Val carriers (see above; e.g. Malhotra et al., 2002; Rosa et al., 2004) agrees well with findings of reduced uncertainty-driven exploration in both groups (Frank et al., 2009; Strauss et al., 2011) and supports the view that prefrontal DA is tightly involved in controlling directed exploration.

To conclude, several human studies have provided evidence for a DA involvement in explore/exploit behavior and related cognitive functions, specifically suggesting a key role for striatal DA in driving reinforcement learning and exploitation, and for prefrontal DA in risk taking and uncertainty-driven exploration.

1.3.3 Limitations of previous studies

The research findings reported above (section 1.3.1 and 1.3.2) should be considered with certain limitations in mind. First of all, a large part of these results are based on genetic association studies, i.e. studies which examine the association between certain DA-related genotypes (e.g. the COMT Val/Met polymorphism) and certain behavioral phenotypes (e.g. directed exploration). However, such genotype-phenotype associations are only correlative and provide, on their own, no evidence that the studied gene is indeed causal for the observed phenotype (see Blanco et al., 2015; Kayser et al., 2015; Li, Tesson, Churchill, & Jansen, 2010). Instead, a third variable, which actually causes the phenotype, might covary with the studied genotype and thereby explain the observed association. For instance, the studied genotype might be "tagged" to another gene variant which is actually causal for the phenotype, an effect called "linkage disequilibrium" (Slatkin, 2008). Furthermore, the studied gene might show different allele frequencies between different subgroups of a population, e.g. different ethnicities, which themselves causally explain the phenotypic variations – an effect called "population stratification" (Hellwege et al., 2017; Tian, Gregersen, & Seldin, 2008). Linkage disequilibrium and population stratification are common limitations of genetic association studies, which also hamper the causal interpretability of the genetic research findings reported above (see Blanco et al., 2015; Li et al.,

2010). Further, many of the reported genetic association studies suffer from relatively small sample sizes, which might increase the risk of yielding unreliable results that cannot be replicated in later studies, as seen for some of the findings reported above (e.g. Barnett et al., 2008; Frank et al., 2009; Kayser et al., 2015). Moreover, another limitation has to be noted for the genetic association studies based on the COMT Val/Met polymorphism, since the COMT enzyme degrades not only DA, but also other catecholamines like epinephrine and norepinephrine. Thus, the COMT Val/Met polymorphism might not only influence DA function, but also other neurotransmitter systems, challenging the interpretation that the observed COMT genotype-phenotype associations clearly represent DA-specific effects. Finally, one should keep in mind that the reported genetic association studies mostly focused on the effect of one single polymorphism, whereas complex phenotypes like explore/exploit behavior surely rely on many more (genetic) influences and their diverse interactions.

Aside from genetic studies, other studies have used the sEBR as a proxy for central DA function, which also brings some limitations. Above all, the validity of the sEBR as a positive predictor of DA function is controversially discussed in the literature, particularly since recent human PET studies found either no or even a negative correlation between the sEBR and different aspects of central DA function (Dang et al., 2017; Sescousse et al., 2018). Furthermore, it remains uncertain which specific aspect of DA function the sEBR reflects, hence most studies simply refer to it as a marker of "DA function" or "DA activity" (see Sescousse et al., 2018). Evidence across different studies remains rather inconclusive about this question, suggesting that the sEBR may (positively or negatively) predict different aspects of DA function, including striatal DA synthesis capacity (Sescousse et al., 2018), D2 receptor availability (Groman et al., 2014), and D1 receptor activity (e.g. Kotani et al., 2016; see also review by Jongkees & Colzato, 2016). This uncertainty makes it difficult to mechanistically interpret the findings of sEBR studies in the light of DA function. In addition, there is a large variability in the methods applied to measure the sEBR, as well as in the measurement conditions, durations, and time points (Jongkees & Colzato, 2016). This methodological heterogeneity may cause a variability in sEBR data not attributable to DA function that further complicates the interpretability of these data within and across studies. Lastly, many sEBR studies, including the ones reported above (see 1.3.2; Dreisbach et al., 2005; Müller et al., 2007; Sherman & Wilson, 2016; Tharp & Pickering, 2011), only analyze sEBR data under the assumption of a linear relationship between the sEBR and DA-related functions, e.g. by a linear correlation/regression analysis or by median splitting into low and high sEBR groups (see Jongkees & Colzato, 2016). However, such approaches do not account for a potential nonlinear (e.g. quadratic) relationship in the data, as suggested by the inverted-U hypothesis of DA (see 1.2.4; e.g. Cools & D'Esposito, 2011). Indeed, some studies found performance to be optimal at an intermediate sEBR, while both lower and higher sEBRs were associated with reduced performance (e.g. Akbari Chermahini & Hommel, 2010, 2012; Dang et al., 2016; Ueda et al., 2016), a pattern which might be overlooked when only testing for linear associations. Thus, although there are good reasons for using the sEBR as

a cheap and noninvasive marker of DA function (see 1.2.5), these limitations should be kept in mind when drawing conclusions about DA effects based on sEBR data.

Another limitation concerns studies based on patients with DA-related diseases like PD and schizophrenia. Both these diseases involve very heterogeneous phenotypes resulting from a complex pathomechanism that is not reducible to a DA dysfunction (see Barone, 2010; Galvan & Wichmann, 2008; Sawa & Snyder, 2002; Tost, Alam, & Meyer-Lindenberg, 2010; Tsuang, 2000). For instance, previous studies have shown that also changes in serotonergic, glutamatergic, and GABAergic neurotransmission are involved in the neuropathology of both PD (Barone, 2010; Bonnet, 2000) and schizophrenia (Brisch et al., 2014; Gill & Grace, 2016; Sawa & Snyder, 2002). Thus, caution is warranted when concluding DA-specific effects from these patient studies without accounting for the potential involvement of other neurotransmitter systems. Additionally, the interpretation of these data in terms of DA function is further complicated in studies examining patients already treated with different drugs or drug combinations, as it is mostly the case (e.g. Egan et al., 2001; Frank et al., 2004; Moustafa et al., 2008; Strauss et al., 2011; van Eimeren et al., 2009). These drugs might unspecifically act on, or interact with, different non-dopaminergic neurotransmitter systems (see Stępnicki et al., 2018), making it even more difficult to clearly attribute observed differences between patients and controls to DA-specific functions.

Aside from these study-specific limitations, two general limitations are shared by most of the reported studies. A first basic limitation to these studies is that most of them lack direct evidence on the neural mechanisms underlying the observed effects on explore/exploit behavior. Such evidence might be provided by *in vivo* brain imaging techniques like fMRI or PET, especially when combining these techniques with cognitive modeling (see 2.8.1). Note that while fMRI cannot directly distinguish between activities of different neurotransmitter systems, it nevertheless allows to locate the neural correlates of different DA-dependent behaviors and DA drug effects, thereby facilitating the mechanistic interpretation of these findings (e.g. striatal vs. prefrontal effects). Yet, none of the reported studies applied PET, and only three of them used fMRI, albeit none of them in an actual explore/exploit paradigm (see Egan et al., 2001; Kohno et al., 2016; van Eimeren et al., 2009). Alternatively to fMRI and PET, neuromechanistic interpretations are also facilitated by the use of neural network models which simulate observed behaviors, as reported above (see 1.3.1; Humphries et al., 2012; Mandali et al., 2015). Yet, a limitation to these simulation studies is that their neuronal models were limited to the basal ganglia network, leaving out influences from prefrontal and other cortical regions that are most likely involved in explore/exploit behavior, as shown by previous research (see 1.1.4; e.g. Daw et al., 2006; Raja Beharelle et al., 2015; Zajkowski et al., 2017). Another limitation to these studies is that neuromechanistical hypotheses generated *in silico* from computational models still need to be empirically validated in the living organism by *in vivo* neuroimaging techniques. Without more research providing direct evidence on the neural processes

underlying explore/exploit behavior, mechanistic interpretations from simulation studies attributing this trade-off to DA-specific functions can only remain speculative.

A second basic limitation shared by most of the reported studies is the lack of evidence for a *causal* role of DA in explore/exploit behavior. Such evidence might be provided, for example, from transgenic animal experiments targeting certain DA-specific genes, as it was done in the study of Beeler et al. (2010). It should be kept in mind, however, that the genotype of these transgenic animals is already modified in the embryo and may therefore affect the adult phenotype in complex and unpredictable ways, given the various gene-gene and gene-environment interactions throughout ontogeny. Accordingly, the hyperdopaminergic DAT knock down mice examined by Beeler et al. (2010) might have exhibited (unobserved) differences in their brain development and neuronal organization compared to control mice, which makes it difficult to specifically attribute behavioral changes in these animals strictly to DAT-specific effects. Alternatively, evidence for a causal role of DA in explore/exploit behavior may be provided by randomized, placebo-controlled pharmacological experiments in animals or humans using DA-specific drugs, preferably in combination with *in vivo* neuroimaging techniques to facilitate mechanistic interpretations. Yet, only four of the reported human studies used a pharmacological approach, of which only one was a placebo-controlled experiment in healthy subjects (Kayser et al., 2015), while the other three were on/off medication studies in PD patients (Frank et al., 2004; Moustafa et al., 2008; van Eimeren et al., 2009). The placebo-controlled experiment was based on the COMT inhibitor tolcapone, which might however be less DA-specific than other available DA drugs, given that the COMT enzyme also metabolizes other neurotransmitters than DA (see Männistö & Kaakkola, 1999). Also, this study considered tolcapone effects only on the behavioral level and did not apply neuroimaging. In addition, most of these pharmacological studies failed to account for individual differences in baseline DA levels, which are believed to strongly influence DA drug effects (see 1.2.4). This might represent an important limitation to these studies, since overseeing such a potential baseline dependency may further complicate the interpretation of their results and even render conclusions drawn from group-level analyses misleading. Note, finally, that while there are further neuroscientific approaches allowing for causal inference, like non-invasive brain stimulation (Parkin et al., 2015) or lesion studies (e.g. Kovach et al., 2012), these methods do not provide, by themselves, direct evidence for a dopaminergic basis of the causal factors.

Taken together, while evidence from several animal and human studies suggests that DA may be tightly involved in regulating the explore/exploit trade-off, most of these studies share important limitations that make it difficult to draw concrete mechanistic or causal conclusions from their findings. Further research should aim to overcome these limitations and employ well-controlled experimental study designs in combination with functional neuroimaging to directly test for the causal effects of DA modulation on explore/exploit behavior and its neural correlates.

1.4 The current project

1.4.1 Objectives

While a growing body of evidence suggests that DA may be causally involved in regulating the explore/exploit trade-off, direct evidence in this regard is very limited and still lacking in humans (see 1.3.3). Therefore, the main aim of this study was to experimentally show that DA is causally involved in controlling human explore/exploit behavior and to reveal the neural underpinnings of this DA-dependent control. A second aim was to examine how individual differences in baseline DA function modulate DA drug effects on explore/exploit behavior, and specifically to test whether such an influence follows the inverted-U function of DA (see 1.2.4). Lastly, one additional aim was to advance cognitive modeling in this field by comparing different established and novel cognitive models of explore/exploit behavior and to select the best-fitting model to be applied to the behavioral and neuroimaging data of this study.

1.4.2 Study design and hypotheses

To directly test for DA effects on explore/exploit behavior and its neural correlates, this study used a pharmacological fMRI approach with a double-blind, placebo-controlled, counterbalanced, within-subjects design. According to this design, healthy subjects (n=31) performed the restless four-armed bandit task in the fMRI scanner under three different drug conditions: the DA precursor L-dopa (150 mg), the DA antagonist haloperidol (2 mg), and placebo. While L-dopa is well established to stimulate DA transmission by providing increased substrate for DA synthesis in the brain, haloperidol is known to reduce DA transmission by blocking D2 receptors (see 1.2.3). Choice behavior in the bandit task was analyzed using a hierarchical Bayesian modeling approach (see 2.7). Therefore, different cognitive models of learning and decision making were first compared using Bayesian cross-validation techniques to select the model with highest predictive accuracy for further analysis. Based on the selected model, DA drug effects on explore/exploit behavior and its neural correlates were then examined, the latter by using a model-based trial-by-trial analysis of the fMRI data (see 2.8). Finally, all subjects in this study performed an initial "baseline session", in which different behavioral proxy measures of DA function were assessed (sEBR and WMC; see 1.2.5) to test whether these measures predict individual differences in DA drug effects according to an inverted-U-shaped function (see 1.2.4).

Based on previous research (see 1.3), it was hypothesized that explore/exploit behavior and its neural correlates are modulated by the two DA drugs, L-dopa and haloperidol, compared to placebo. Note that the heterogeneity of findings in this field (see 1.3) and the complexity of the DA brain system and DA drug actions (see 1.2) make it difficult to formulate specific hypotheses about the expected magnitude and direction of the DA drug effects. Still, most research supports the view that an increase in tonic striatal DA promotes random exploration (see 1.3.1; e.g. Beeler et al., 2010), while an increase in prefrontal DA promotes uncertainty-driven exploration (see 1.3.2; e.g. Frank et al., 2009; Kayser et

al., 2015). Therefore, it was hypothesized that both random exploration, as indexed by the softmax β, and uncertainty-driven exploration, as indexed by the exploration bonus parameter φ, are increased under L-dopa vs. placebo and decreased under haloperidol vs. placebo. These behavioral drug effects were expected to be mediated by drug-induced modulations in the activity of brain regions implicated in exploratory choices, foremost the FPC and IPS (see 1.1.4; e.g. Daw et al., 2006). Furthermore, it was expected that DA drug effects on explore/exploit behavior are modulated by the individual DA baseline, as indexed by the sEBR and WMC, according to an inverted-U-shaped function (see 1.2.4; e.g. Cools & D'Esposito, 2011). More specifically, L-dopa should increase exploration most strongly in subjects with below-optimal (depleted) baseline DA levels, while showing no or even a reversed drug effect in subjects with optimal or above-optimal baseline DA levels. In contrast, haloperidol should decrease exploration most strongly in subjects with above-optimal (excessive) baseline DA levels, while showing no or even a reversed drug effect in subjects with optimal or below-optimal baseline DA levels. Yet, note that also alternative hypotheses are possible, since it is unclear how the term "optimal performance" used in the inverted-U hypothesis of DA relates to explore/exploit behavior. Lastly, for the cognitive model comparison, it was expected that choice behavior in the bandit task is best described by a model which accounts for both random and uncertainty-driven exploration, as previous research has shown that humans use both these strategies for exploration (see 1.1.2 and 1.1.3; e.g. Cogliati Dezza et al., 2017; Wilson et al., 2014). Furthermore, it was hypothesized that a novel extension of this model, which also captures reward-independent choice repetition (perseveration), further improves the model fit compared to previous models not accounting for perseveration. Notably, this last hypothesis is based on previous findings showing that humans indeed exhibit perseverative choice behavior, also known as "sticky choice" (Brough, Isaac, & Chernev, 2008; Payzan-LeNestour & Bossaerts, 2012; Rutledge et al., 2009; Schönberg, Daw, Joel, & O'Doherty, 2007; Worthy, Pang, & Byrne, 2013; see also Lau & Glimcher, 2005). Moreover, it has been argued in the literature that if perseveration is not explicitly accounted for in the cognitive model, it might be captured by the exploration bonus parameter as a tendency to avoid uncertain options, making it more difficult to detect uncertainty-driven exploration (see Badre et al., 2012; Payzan-LeNestour & Bossaerts, 2012; further discussed in section 6.5.2).

2 Methods

2.1 Participants

In total, 34 healthy male subjects participated in the study (aged 19 to 35 years, *M* = 26.85, *SD* = 4.01), of which 31 completed all experimental sessions (see 2.2). Three subjects dropped out of the study due to illness or personal reasons, two after the initial baseline session and one after the first fMRI session, but their behavioral data from the baseline session were included in the analysis. Only males were included, owing to the rationale that female menstrual cycle effects are an unwanted source of variance, especially since ovarian hormones have been shown to interact with central DA function (see reviews by Almey, Milner, & Brake, 2015; Yoest, Quigley, & Becker, 2018). Participants were recruited via an online job portal (www.stellenwerk.de) and included mainly students from the University of Hamburg. Subjects who signed up via e-mail or telephone were subsequently screened in a telephone interview for the following inclusion criteria: male, age 18-35 years, normal weight (BMI 18.5-25.0), right-handed, fluent German in speaking and writing, normal or corrected to normal vision, no hearing impairments, no major past or present psychological, neurological, or physical disorders, non-smoker, no excessive consumption of alcohol (< 10 glasses per week), no consumption of illegal drugs, no consumption of prescription drugs within the two months prior to the study, no irreversibly attached metal in or on the body, no claustrophobia (the latter two due to the fMRI measurement). Before participating in the study, all subjects provided informed written consent and completed a medical check by a physician including an electrocardiogram (ECG) and an interview about their medical history and present health status. Only subjects who passed this medical check were allowed to take part in the study. Participants were paid afterwards with a fixed amount (270 €) plus variable monetary rewards from the decision tasks (30-50 €). The study procedures were approved by the local ethics committee (Hamburg Medical Council).

2.2 General procedure

This pharmacological fMRI study employed a double-blind, placebo-controlled, counterbalanced, within-subjects design (see Pine et al., 2010). Each subject was tested in four separate sessions: one baseline session and three fMRI sessions. At the baseline session, participants were invited in groups of four either in the morning (9 am to 1 pm) or in the afternoon (2 pm to 6 pm) for the medical check, the signing of study agreements, and for the behavioral baseline screening (see 2.3). The baseline session was scheduled five to six days prior to the first fMRI session. At the three fMRI sessions, each participant performed two decision tasks in the fMRI scanner under three different drug conditions: the DA precursor L-dopa, the DA antagonist haloperidol, and placebo. The order of drug conditions was counterbalanced across subjects, and both subjects and experimenter were blinded to the drug order. For each subject, fMRI sessions were scheduled exactly one week apart and at the same time of day. Each fMRI session lasted about 4.0 to 4.5 hours, starting between 8 am and 2 pm depending on

the participant. The procedure of each fMRI session was as follows: Upon arrival, subjects first signed fMRI agreements and completed several questionnaires on their physical wellbeing and mood (see 2.9.3). After that, and exactly 2.5 hours before testing in the fMRI scanner started, participants received a first pill containing either 2 mg haloperidol or placebo. Two hours later, i.e. exactly 0.5 hours before testing in the fMRI scanner started, subjects received a second pill containing either Madopar (150 mg L-dopa + 37.5 mg benserazide) or placebo. Placebo pills contained maize starch and were indistinguishable from the drugs. Over the whole experiment, each subject received one dose of Madopar in one session, one dose of haloperidol in another session, and two placebo pills in the remaining session (with counterbalanced drug order, see above). The administration of two pills per session was necessary to account for the different pharmacokinetics of both drugs without revealing the drug order to the participant or experimenter by the time of drug administration. Testing in the fMRI scanner started 0.5 hours after ingestion of the second pill, aiming to achieve a peak plasma concentration of the drug (Madopar or haloperidol) approximately halfway through the one hour testing in the fMRI scanner. In the scanner, subjects first performed the restless four-armed bandit task (see 2.4), followed by a short reinforcement learning task that was not further analyzed in this study. Both tasks were trained on a practice version outside the scanner prior to the fMRI testing. On the first fMRI session, a structural MR image was acquired directly after fMRI testing (see 2.8.2). Each fMRI session ended with a post-fMRI testing outside the scanner (see 2.5), lasting about 30 min, before the participant was released. Throughout each fMRI session, several control variables were assessed at different time points, including physical wellbeing parameters and mood (see 2.9.3). Subjects were not allowed to eat or drink anything but water throughout the fMRI session, but were offered a small snack (cereal bar) after testing in the fMRI scanner to aid concentration for the post-fMRI testing. Data assessment for all subjects and sessions was conducted by the same experimenter with the help of two trained assistants according to fixed protocols.

2.3 Baseline screening

The baseline screening included several components and lasted about two hours on average. The screening was performed in groups of four in the same experimental room, each subject sitting in front of a computer screen, separated by partition walls. An experimenter was present in the room at all times. The screening started with a measurement of the spontaneous eye blink rate (ca. 15 min), followed by a computerized testing of working memory capacity and discounting behavior (ca. 75 min), and ended with several questionnaires (ca. 30 min). Note that the discounting tasks and most of the questionnaires were included for another project and were not further analyzed in this study. The computerized testing included six tasks in a fixed order: (1) Rotation Span Task, (2) Operation Span Task, (3) Delay Discounting Task, (4) Listening Span Task, (5) Digit Span Task, (6) Probability Discounting Task. Participants were encouraged to take small breaks in between the tasks to aid concentration. In the following, all relevant components of the baseline screening will be described in more detail.

2.3.1 Spontaneous eye blink rate (sEBR)

The spontaneous eye blink rate (sEBR) at baseline was measured via electromyography (EMG) for 5 min under resting conditions. For the measurement, three Ag/AgCl electrodes were attached to the participant's face. Two electrodes were placed directly below the left eye, one of them centrically and one peripherally with 2-3 mm distance to the central one (see Blumenthal et al., 2005). A ground electrode was placed in the middle of the forehead. Recording was performed with a sampling rate of 1000 Hz and an online bandpass filter of 28-500 Hz, using an MP100 hardware system with the software AcqKnowledge (version 3.9.1; Biopac Systems, Goleta, CA).

For recording, subjects were instructed to sit in front of a computer screen and look straight ahead at a fixation cross for 5 min. Participants were told to behave as quietly as possible and not to talk, move, or grimace during the recording. Crucially, participants were neither informed about the actual purpose of the measurement, nor were they instructed in any manner about blinking. Instead, they were told that the purpose of the measurement was to record eye muscle activity under rest and that blinking during recording was "not a problem". This cover story was used to avoid that participants became aware of their blinking behavior and changed it deliberately during the measurement, e.g. by trying not to blink or to blink more or less than they would naturally do. Illumination in the testing room was kept constant for all participants. In addition, all data were recorded before 5 pm, since the sEBR is stable during daytime, but increases in the evening (Barbato et al., 2000). EMG recordings were used to extract the number of spontaneous eye blinks per minute for each participant using the software MATLAB (R2014b; MathWorks, Natick, MA). For this, the 5 min recordings were divided into time windows of 10 s. Within each window, the number of peaks exceeding a locally determined threshold was counted using the MATLAB function *findpeaks*. The threshold for each time window was set to the local mean of the data plus four times the local standard deviation. The total number of peaks across all time windows was then divided by five to yield the sEBR per minute.

2.3.2 Working memory capacity (WMC)

Working memory capacity (WMC) was measured using four different tasks: the Rotation Span Task, Operation Span Task, Listening Span Task, and Digit Span Task. All four tasks were implemented using the software MATLAB (R2014b; MathWorks, Natick, MA) with the Psychophysics Toolbox extensions (version 3.0.12; Brainard, 1997; Kleiner et al., 2007).

The Rotation Span Task was adopted from Foster et al. (2015) and belongs to the class of complex span tasks for measuring WMC (see Kane et al., 2004; Redick et al., 2012; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). In this task, subjects were required to memorize a sequence of different arrows (memory component) while being distracted by a letter rotation task (distractor component). The memory component of the task is described first. In each trial, a random series of two to five arrows out of 16 different arrows (eight short and eight long arrows pointing in eight different radial directions) was presented on screen. Each arrow appeared on screen for 650 ms, followed by a blank

screen for 250 ms. Thereafter, subjects had to recall all arrows of the preceding series in the same order as they were presented. For recall, all 16 possible arrows were displayed on screen and subjects were asked to click on the arrows that appeared in the preceding series in the correct order. Participants had no response time limit for recall. At the end of each trial, feedback about the number of correctly recalled arrows was presented on screen. In addition to this memory component, each trial also contained a distractor component. Before the presentation of each arrow within a trial, subjects had to solve an item of a letter rotation task. Each trial therefore contained as many letter rotation items as arrows (i.e. two to five), both being presented intermixed within one trial. In each item of the letter rotation task, one of four letters (F, G, J, or R) rotated at different angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°) appeared on screen, being either mirror-inverted or not. The task was to mentally rotate the letter and to indicate whether it was mirror-inverted or not. About half of all letters presented in the task were mirror-inverted. The response time limit for each subject was determined from their mean reaction time plus 2.5 standard deviations in the preceding practice block of the letter rotation task (see below). If participants exceeded this limit, the letter disappeared and the item was counted as an error. After each item of the letter rotation task, a blank screen appeared for 200 ms before the next arrow was shown. At the end of each trial, subjects received feedback about their performance in the letter rotation task (percent accuracy), averaged over all preceding trials. Participants were instructed to always maintain accuracy levels in the letter rotation task above 85 %. In total, the Rotation Span Task contained 12 trials, including three trials of each set size (two to five) in random order. Thus, 42 arrows and 42 letters were presented in total. At the end of the task, two different memory scores were calculated: the partial score and the absolute score. The partial score equals the number of correctly recalled arrows out of all 42 presented arrows. The absolute score only includes the number of correctly recalled arrows within sets in which all arrows were recalled correctly. In preparation for the task, participants completed three practice blocks. In the first practice block, only the memory component of the task was trained for four trials (two trials of set size two and two trials of set size three). In the second practice block, only the distractor component of the task was trained for 15 trials. After this block, the mean reaction time plus 2.5 standard deviations was calculated for each subject to determine the response time limit for this subject in the distractor component of the final task (see above). The third practice block trained the final task with both its memory and distractor component and contained three trials of set size two.

The Operation Span Task was adopted from Foster et al. (2015) and also belongs to the class of complex span tasks for measuring WMC (see above). The general procedure of this task was very similar to the Rotation Span Task described above, except for the type of stimuli involved. In this task, subjects were required to memorize a sequence of letters (memory component) while being distracted by math operations to be solved (distraction component). The memory component of the task is described first. In each trial, a random series of three to seven letters out of 12 different letters (F, H, J, K, L, N, P, Q, R, S, T, Y) was presented on screen. Each letter appeared on screen for 1 s, followed by a blank screen

for 250 ms. Thereafter, subjects had to recall all letters of the preceding series in the same order as they were presented. For recall, all 12 possible letters were displayed on screen and subjects were asked to click on the letters that appeared in the preceding series in the correct order. Participants had no response time limit for recall. At the end of each trial, feedback about the number of correctly recalled letters was presented on screen. In addition to this memory component, each trial also contained a distractor component. Before the presentation of each letter within a trial, subjects had to solve an item of a math operation task. Each trial therefore contained as many math operation items as letters (i.e. three to seven), both being presented intermixed within one trial. In each item of the distractor task, a math problem (e.g. (1*2) + 1 = ?) appeared on screen. Subjects were required to solve this problem as quickly as possible and press a button as soon as they had a solution. A potential solution was then displayed on screen (e.g. 3) and participants had to indicate whether this solution was correct or false. The response time limit for each subject was determined from their mean reaction time plus 2.5 standard deviations in the preceding practice block of the math operation task (see below). If participants exceeded this limit, the math problem disappeared and the item was counted as an error. After each item of the math operation task, a blank screen appeared for 200 ms before the next letter was shown. At the end of each trial, subjects received feedback about their performance in the math operation task (percent accuracy), averaged over all preceding trials. Participants were instructed to always maintain accuracy levels in the math operation task above 85 %. In total, the Operation Span Task contained 15 trials, including three trials of each set size (three to seven) in random order. Thus, 75 letters and 75 math operations were presented in total. At the end of the task, two different memory scores were calculated: the partial score and the absolute score. The partial score equals the number of correctly recalled letters out of all 75 presented letters. The absolute score only includes the number of correctly recalled letters within sets in which all letters were recalled correctly. In preparation for the task, participants completed three practice blocks. In the first practice block, only the memory component of the task was trained for four trials (two trials of set size two and two trials of set size three). In the second practice block, only the distractor component of the task was trained for 16 trials. After this block, the mean reaction time plus 2.5 standard deviations was calculated for each subject to determine the response time limit for this subject in the distractor component of the final task (see above). The third practice block trained the final task with both its memory and distractor component and contained three trials of set size two.

The Listening Span Task was adapted from the German version of the automated Reading Span Test developed by van den Noort, Bosch, Haverkort, and Hugdahl (2008), which is based on the original task by Daneman and Carpenter (1980). In the Reading Span Test, subjects are required to read a series of sentences aloud and to recall the last word of each sentence directly after. For the baseline screening, a listening version of this test was developed in order to run the task completely computerized on multiple subjects simultaneously. In the listening version of the task, subjects also needed to recall the final word of each sentence, but here they listened to the list of sentences rather than reading it aloud

(similar to the Listening Span Test also developed by Daneman & Carpenter, 1980). For recall, subjects typed the remembered words into the computer rather than saying them aloud to an experimenter. Except for these changes, the Listening Span Task used exactly the same sentences and task structure as the Reading Span Test by van den Noort et al. (2008). The task contained five blocks, each block including one trial of each set size, with set sizes ranging from two to six sentences. Thus, 100 sentences were presented in total. Within each block, the order of set sizes was the same for each subject and taken from van den Noort et al. (2008), who generated a random order of set sizes for each block. The block structure was introduced to distribute trials of different set sizes more evenly throughout the task, but was not visible to the subjects in any form. In each trial, subjects listened to a sequence of recorded German sentences over headphones and had to memorize each sentence's final word. The length of the sentences was controlled for, ranging from 12 to 17 words. After each sentence, there was a time interval of 1 s before the next sentence started. After the last sentence of the series, the German word "Erinnern" (recall) was displayed on screen along with a text box, in which participants could type in the words they recalled. The order of recall was free and there was no time limit for recall. All words to be recalled were simple and frequent German words like "Park" (park), "Kaffee" (coffee), or "Anfang" (beginning). After recall, subjects pressed a button to proceed to the next trial. During the task, participants received no feedback about their memory performance. At the end of the task, the total number of correctly recalled words over all trials was determined (span score). In preparation for the task, subjects performed two practice trials (of set size two and three), in which they received feedback about the number of correctly recalled words.

The Digit Span Task was adopted from the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008) and contained a forward and a backward version. In this task, subjects heard a series of numerical digits and had to recall the series in the normal order (forward version) or in the reverse order (backward version) directly after. For the baseline screening, the task was adapted for PC in order to test multiple participants simultaneously. Here, the digits were presented over headphones and subjects typed the recalled digits into the computer rather than saying them aloud. Except for these changes, the computerized version of the Digit Span Task used exactly the same task structure and digit sequences as the original task by Wechsler (2008). In the forward version, set sizes ranged from three to eight digits and in the backward version from two to seven. Both versions contained two trials per set size, whereby set sizes were presented in ascending order. Each trial started with the presentation of a digit sequence with a speed of one digit per second. After that, a question mark appeared on screen to prompt subjects to type in the digits they recalled in the correct order. There was no time limit for recall and subjects received no feedback about their memory performance. Participants needed to recall at least one trial per set size correctly in order to proceed to the next larger set size. The task terminated if both trials of the same set size were not recalled correctly. Thus, the number of trials depended on the subject's performance with a maximum of 12 trials for both the forward and the backward version. At the end of the task, two different types of scores were calculated

for each version: the span score and the total score. The span score is the size of the longest digit sequence recalled correctly in the normal order (span score forward) or in the reverse order (span score backward). The total score equals the total number of correct trials, calculated separately for the forward version (total score forward) and for the backward version (total score backward). The maximum achievable scores were eight for the span score forward, seven for the span score backward, and 12 for the total scores of both versions. In preparation for the task, participants performed one practice trial of set size three for each version, in which they received feedback about the correctness of their answer.

2.3.3 Questionnaires

At the end of the baseline screening, subjects completed a computer-based survey including several questionnaires assessing demographics, personality traits, addictive behavior, and various symptoms of psychopathology. The questionnaires were presented in a fixed order and were all implemented using the online survey application LimeSurvey (LimeSurvey GmbH, Hamburg, Germany). Since most of these questionnaires were assessed for a different study, methods and results of the survey are not further reported here. Only two questionnaires were included in the data analysis. First, data of the Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1992; German version by Franke, 1995) were analyzed to ensure that subjects included in this study did not exhibit any severe psychiatric symptoms. For this, mean scores for each of the nine subscales (somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism) as well as the Global Severity Index (GSI), which is the mean value of all 90 items, were calculated for each subject and transformed into T values based on a German norm sample of male students (see SCL-90-R manual by Franke, 2000, p. 310-329). As instructed in this manual, the screening cut-off was set to $T_{GSI} \geq 63$ or $T \geq 63$ for at least two of the nine subscales, which was reached by none of the participants. Second, data of the Edinburgh Handedness Inventory (EHI; Oldfield, 1971) were analyzed to ensure that all participants were right-handed. Results confirmed this, as laterality quotients ranged between 40 and 100 ($M = 91.71$, $SD = 13.29$), with positive scores indicating the dominance of a person's right hand in everyday activities.

2.4 Bandit task

The restless four-armed bandit task was adapted from Daw et al. (2006). The task included 300 trials, which were separated by short breaks into four blocks à 75 trials. Each trial started with the presentation of four different colored squares ("bandits") representing four choice options (see Figure 4a). The squares were displayed on a screen that was reflected in a head coil mirror inside the fMRI scanner (see 2.8.2). Subjects selected one option using a button box held in their right hand, which had four buttons in the same relative positions as the four bandits on screen. Subjects had a maximum of 1.5 s to indicate their choice. If no button was pressed during that time, a large red X was displayed for

4.2 s in the center of the screen indicating a missed trial with no points earned. If subjects pressed a button on time, the selected bandit was highlighted by a black frame. After a waiting period of 3 s, signaled by the subsequent presentation of three black dots within the chosen bandit, the number of points earned in this trial was displayed within this bandit for 1 s. After that, the four bandits disappeared and only a fixation cross remained on screen until the trial ended 6 s after trial onset, followed by a jittered intertrial interval before the new trial started. Durations of the intertrial intervals followed a Poisson distribution with a mean of 2 s, thereby mostly ranging between 0-5 s. At the end of the task, the sum of all points earned as well as the monetary payout resulting from these points were displayed on screen. Participants were told in advance that 5 % of all points earned would be paid out as cents after the experiment (i.e. 5 cents per 100 points).

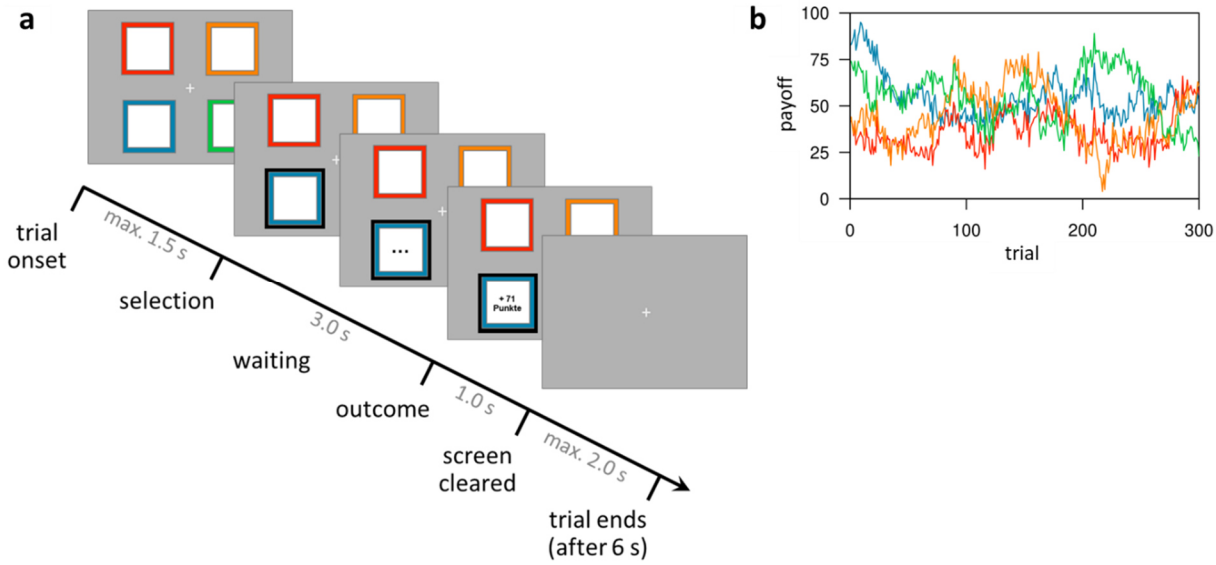


Figure 4. Task design of the restless four-armed bandit task. (**a**) Illustration of the timeline within a trial. At trial onset, four colored squares (bandits) are presented. The subject selects one bandit within 1.5 s, which is then highlighted and, after a waiting period of 3 s, reveals its payoff for 1 s. After that, the screen is cleared and the next trial starts after a fixed trial length of 6 s plus a variable intertrial interval (not shown) with a mean of 2 s. (**b**) Example of the underlying reward structure. Each colored line shows the payoffs of one bandit (mean payoff plus Gaussian noise) that would be received by choosing that bandit on each trial. See text for detailed descriptions. Both figures adapted from Daw et al. (2006).

The mean payoffs of the four bandits drifted randomly across trials according to a decaying Gaussian random walk (as described by Daw et al., 2006). At each trial t, the payoff of bandit i was drawn from a Gaussian distribution with mean value $\mu_{i,t}$ and variance $\sigma_o^2 = 4^2$ (observation variance) and rounded to the nearest integer between 1 and 100. From one trial to the next, the mean value of bandit i changed according to $\mu_{i,t+1} = \lambda\mu_{i,t} + (1-\lambda)\vartheta + v_t$, with a decay parameter $\lambda = 0.9836$, decay center $\vartheta = 50$, and diffusion noise v drawn independently in each trial from a Gaussian distribution with zero mean and variance $\sigma_d^2 = 2.8^2$ (diffusion variance). Daw et al. (2006) generated three instantiations of this process for their bandit task, which were also used in the three fMRI sessions of the current study. One of these instantiations is shown in Figure 4b. Each colored line reflects the payoffs of one bandit

(mean payoff plus observation variance) that would be received by choosing that bandit on each trial. The order of these three instantiations across fMRI sessions was the same for all subjects, thereby unconfounded with the drug order, which was counterbalanced across subjects. The task was implemented using the software MATLAB (R2014b; MathWorks, Natick, MA) with the Psychophysics Toolbox extensions (version 3.0.12; Brainard, 1997; Kleiner et al., 2007).

2.5 Post-fMRI testing

After testing in the fMRI scanner, subjects completed a post-fMRI testing to assess different control variables in order to test for DA drug effects on these variables. On the one hand, this testing included three tasks from a test battery of attentional performance (see below) and the Digit Span Task (forward and backward). These measures were included as control variables to test whether DA drug effects on explore/exploit behavior might be mediated by drug-induced alterations in attention or working memory. On the other hand, also the spontaneous eye blink rate (sEBR) was assessed in this post-fMRI testing to test if this measure would be sensitive to the DA drug condition, as partly suggested by previous research (reviewed by Jongkees & Colzato, 2016; although see Mohr, Sándor, Landis, Fathi, & Brugger, 2005). The testing was computer-based and conducted individually in a quiet room next to the fMRI scanner with an experimenter being present in the room. In the following, all components of the post-fMRI testing will be described in more detail.

The Tests of Attentional Performance (TAP, version 2.3; Zimmermann & Fimm, 2012) is a computerized test battery that measures different aspects of attentional performance. Three subtests of the TAP were used in the post-fMRI testing: Alertness, Go/NoGo, and Flexibility. All tasks were implemented using the TAP 2.3.1 software and the corresponding response buttons delivered with this software.

The subtest Alertness measures the general wakefulness of a subject, which enables the subject to respond quickly and accurately to a given demand (Zimmermann & Fimm, 2012). The subtest included four blocks, in which reaction times were assessed under two conditions. In two blocks (first condition), subjects needed to respond as quickly as possible to a white cross (+) appearing in randomly varying time intervals in the center of a black screen. Responses were made by button press using the index finger of the dominant (right) hand. These two blocks measured intrinsic alertness. The other two blocks (second condition) were similar, but a warning tone (cue) preceded the presentation of the white cross. Subjects were instructed to only respond to the white cross, but not to the warning tone. These two blocks measured phasic arousal. In each block, the white cross was presented 20 times. The order of the blocks was fixed according to an ABBA design (A = without warning tone, B = with warning tone) to compensate for effects of fatigue. In total, the subtest took about 5 min. Reaction times for correct responses and number of errors (misses) were examined, as well as the index of phasic arousal, calculated as the difference of the median reaction time in condition A minus B, divided by the median reaction time across both conditions (see Zimmermann & Fimm, 2012).

The subtest Go/NoGo measures a subject's ability to perform appropriate responses quickly and accurately while simultaneously suppressing inappropriate responses. This ability for selective reaction is considered an important aspect of behavioral control (see Zimmermann & Fimm, 2012). The subtest contained one block, in which two types of stimuli were shown successively in the center of the screen in randomly varying time intervals. Subjects were instructed to respond as quickly as possible to the upright cross (+), but not to the diagonal cross (×). Both stimulus types were presented 20 times in random order. This subtest took about 2 min. Reaction times for correct responses and number of errors (misses and false alarms) were examined.

The subtest Flexibility is a set shifting task that measures a subject's ability to actively switch attention between different tasks or objects (see Monsell, 2003; Zimmermann & Fimm, 2012). This subtest contained one block of 100 trials. In each trial, two types of stimuli (one letter and one number) were presented simultaneously on the left and right side of the screen. Subjects needed to indicate as quickly as possible at which side of the screen a target stimulus appeared. Crucially, the type of the target stimulus changed from trial to trial, alternating between letters and numbers. In the first trial, subjects had to indicate the side of the letter, in the next trial the side of the number, and so on. Responses were made by pressing either a left button or a right button with the index finger of the corresponding hand. This subtest took about 4 min. Reaction times for correct responses and number of errors (incorrect response side) were examined.

The Digit Span Task followed the same procedure as in the baseline screening (see 2.3.2). For each of the three fMRI sessions, a new stimulus set of random digit sequences was generated for both the forward and backward version of the task, which differed from the set presented in the baseline screening. The order of these three stimulus sets across fMRI sessions was the same for all subjects and thus unconfounded with the drug order, which was counterbalanced across subjects.

The sEBR measurement in the post-fMRI testing followed the same procedure as in the baseline screening (see 2.3.1), except for one difference: Here, the sEBR was assessed using a video-based procedure instead of electromyography (EMG) due to the lack of an EMG equipment in the testing room. Therefore, subjects were video-recorded while they looked at a central fixation cross on the screen for 5 min under resting conditions. Instructions for the measurement were exactly the same as in the baseline screening. Additionally, subjects were informed about the video recording and agreed to this procedure before the measurement started. Illumination in the testing room was kept constant for all subjects and testing sessions. After all sessions, video records were analyzed by a trained assistant (blinded to the drug condition), who counted the number of eye blinks that occurred during the 5 min interval. The total number of eye blinks was divided by five to yield the sEBR per minute.

2.6 Additional control variables

In addition to the first set of control variables assessed in the post-fMRI testing, another set of control variables was assessed at several time points throughout each fMRI session and included different measures of subjective mood and physical wellbeing. While the latter were mainly assessed to monitor subjects' wellbeing throughout the testing, all control variables in this set were later tested for DA drug effects in order to rule out these factors as potential mediators of DA drug effects on explore/exploit behavior.

First, subjective mood was assessed with two paper-and-pencil instruments at three different time points: (1) before ingestion of the first pill, (2) directly before testing in the fMRI scanner, and (3) directly after testing in the fMRI scanner. The first instrument was a visual analog scale (VAS; Bond & Lader, 1974) that measured subjective feelings on 16 bipolar dimensions, such as "happy – sad", "interested – bored", and "strong – feeble". For each dimension, subjects marked a position along a horizontal line (100 mm) that indicated how they felt at that moment in relation to the two extremes at both line ends. For further analysis, scores on these 16 dimensions were summarized into three subscale scores called "alertness", "contentedness", and "calmness", as described by Bond and Lader (1974). In short, after reversing some of the items, all items were log-transformed and grouped into three subscales, then subscale scores were calculated by a weighted summation over all items belonging to that subscale, weighting each item by its respective factor loading. The second instrument was a pictorial rating system called Self-Assessment Manikin (SAM; Lang, 1980; see also Bradley & Lang, 1994), which examined subjective states on three dimensions. The three dimensions were "pleasure" (i.e. feeling happy or unhappy), "arousal" (i.e. feeling calm or excited), and "dominance" (i.e. feeling in control or controlled). Each dimension was presented by a series of five pictograms showing figures in different emotional states, which marked different points along a nine-point rating scale.

Moreover, physical wellbeing of participants was examined at four different time points: (1) before ingestion of the first pill, (2) one hour after ingestion of the first pill, (3) directly before testing in the fMRI scanner, and (4) at the very end of the post-fMRI testing, i.e. immediately before the participant was released. First, vital parameters, including pulse and blood pressure (systole and diastole), were measured by the experimenter or a trained assistant. In addition, a paper-and-pencil questionnaire was used to assess ten potential drug side effects, including vertigo, nausea, blurred vision, headache, tremor, irregular heartbeat, lethargy, inner unrest, dry mouth, and dry skin. Each side effect was measured on a seven-point rating scale ranging from 0 (not present) to 6 (extreme).

At the end of each fMRI testing day, subjects were asked to guess which drug they had received at that day, as well as how confident they were about that guess on a five-point rating scale ranging from 1 (very uncertain) to 5 (very certain). Note that subjects were instructed to make their drug guesses

on each fMRI session independently of the other sessions, meaning that they were also allowed to guess the same drug more than once.

2.7 Cognitive modeling

2.7.1 Introduction to hierarchical Bayesian modeling

The theoretical background of hierarchical Bayesian modeling is only briefly described in this section. For a more comprehensive introduction to this method, the reader is referred to the textbooks of Kruschke (2015), Lee and Wagenmakers (2015), and Gelman (2014).

Bayesian modeling is based on the Bayes rule for conditional probabilities, which states that the probability of an event A, given that another event B has been observed, is:

$$P(A|B) = P(B|A)\,P(A)\,/\,P(B) \quad (with\ P(B) \neq 0).$$

Herein, $P(A|B)$ is the conditional probability of A given B, $P(B|A)$ the conditional probability of B given A, and $P(A)$ and $P(B)$ are the unconditional probabilities of event A or B, respectively. In Bayesian cognitive modeling, this rule is used to estimate the parameter values of a cognitive model from the observed data and prior knowledge. Analogous to the formula above, the probability (for a discrete parameter) or density (for a continuous parameter) of a parameter value (θ) given the observed data (D) can be expressed as:

$$p(\theta|D) = p(D|\theta)\,p(\theta)\,/\,p(D) \quad (with\ p(D) \neq 0).$$

Herein, $p(\theta|D)$ is called the posterior distribution of the parameter, $p(D|\theta)$ the likelihood, and $p(\theta)$ the prior distribution, or simply prior. The term $p(D)$, called the marginal likelihood, only serves as a normalizing constant, which is usually of minor importance in Bayesian modeling as it does not affect the relative posterior probabilities of different parameter values. Thus, the formula above simplifies to:

$$p(\theta|D) \propto p(D|\theta)\,p(\theta),$$

which states that the posterior distribution is proportional to the likelihood times the prior. One important aspect in Bayesian inference is that both the prior belief about the parameter (before observing the data) as well as the posterior belief (after observing the data) are expressed as probability distributions, which assign each possible parameter value a relative probability of being true. For example, if no prior knowledge about the parameter exists, this might be expressed by a uniform prior distribution that assigns equal probability to each possible parameter value. However, if prior knowledge, e.g. gained by previous studies, renders certain parameter values more probable, this can be expressed by assigning these values a higher prior probability compared to values that are less likely based on prior knowledge. Accordingly, the posterior distribution expresses the relative probability of each parameter value after observation of the data, which thereby also reflects the degree of uncertainty about the true parameter value. Note that this approach clearly distinguishes

Bayesian parameter estimation from frequentist approaches like maximum likelihood estimation (MLE), which do not account for prior knowledge and only provide classical point estimates (or interval estimates) for the true parameter value. Finally, in order to arrive at this posterior distribution, the prior needs to be combined with the likelihood, which contains information from the observed data. The likelihood function describes the probability of the observed data given the parameter value(s), which can be determined from the cognitive model. For example, a reinforcement learning model (see 1.1.3) allows to calculate for each set of parameter values (e.g. learning rate and softmax parameter) the choice probabilities of all actions that have actually been observed, and thereby the overall probability of these data given the parameter values (likelihood).

As the posterior distribution is typically a complex function which cannot be solved analytically, it is usually numerically approximated by a class of sampling methods called Markov Chain Monte Carlo (MCMC; Robert & Casella, 2005). Commonly used MCMC algorithms in Bayesian modeling include the Metropolis-Hastings algorithm, the Gibbs sampler (e.g. in the software WinBUGS and JAGS), and the Hamiltonian Monte Carlo algorithm (e.g. in the software Stan; see 2.7.2). The basic idea behind these MCMC algorithms is to generate a random number sequence ("Markov chain") through parameter space that "visits" parameter values with higher posterior probability more often, thereby yielding a representative sample from the posterior distribution once the chain has reached equilibrium. In other words, the stationary distribution of the Markov chain equals the posterior distribution. To ensure that the stationary distribution has been reached, users usually run multiple independent Markov chains for several thousand steps, discard their initial "burn-in" period, and check their samples after burn-in for convergence by statistical tests. For instance, one such test for convergence is the so called $\hat{R}$ ("R hat") statistic (Gelman & Rubin, 1992), which basically tests if the ratio of the between-chain to the within-chain variance is close to one, whereas values above 1.1 indicate inadequate convergence.

Hierarchical Bayesian modeling combines the Bayesian modeling approach with the use of hierarchical models to describe the data (see Gelman & Hill, 2007; Kruschke & Vanpaemel, 2015). These models involve multiple parameter levels to reflect hierarchical dependencies within the data, e.g. between single subjects belonging to a group. To express these dependencies, the model assumes that subject-level parameters describing individual behavior are drawn from a higher-level distribution specified by group-level parameters, also called hyperparameters. This higher-level distribution is often modeled as a Gaussian, whose mean and variance are then hyperparameters of the model. The subject-level and group-level parameters form a joint parameter space and are estimated simultaneously, whereby data from different subjects mutually inform each other via the higher-level group parameters. As a result, hierarchical models show "shrinkage" of the subject-level parameters towards the group-level mean (Efron & Morris, 1977; Lehmann & Casella, 1998). The degree of shrinkage depends on the estimated group-level variance, which is in turn informed by the actual between-subject variance in the data. By pulling extreme values towards more plausible values, shrinkage serves to reduce the impact of sampling noise in the data (see Gelman, Hill, & Yajima, 2012; Kruschke, 2013; Kruschke &

Vanpaemel, 2015). Moreover, since each subject-level parameter is informed by data of the entire group, hierarchical modeling often provides stable parameter estimates even with sparse data, e.g. few trials per subject (Ahn, Krawitz, Kim, Busmeyer, & Brown, 2011; Katahira, 2016). Another benefit of hierarchical models is that hyperparameters can be meaningfully interpreted to reflect overall group tendencies and can be used to directly compare different subpopulations (e.g. healthy vs. atypical) or different experimental conditions with each other, as done in the current study (see 2.7.2).

2.7.2 Cognitive modeling in the current study

Choice behavior in the four-armed bandit task was modeled using several cognitive models of learning and decision making in order to compare these models and select the best one (in terms of predictive accuracy) for further analyses. Each of the applied cognitive models was composed of two components: First, a learning rule describing how participants update subjective value estimates for each choice option (bandit) based on previous choices and obtained rewards. Second, a choice rule modeling how these learned value estimates influence future choices. By combining two different learning rules with three different choice rules, a total of six cognitive models resulted for model comparison. In the following, all cognitive models are introduced first, before the procedure of their parameter estimation is described.

The first learning rule was the "Delta rule" (Sutton & Barto, 1998), which is an established temporal difference model of reinforcement learning. According to this rule, subjects update the expected reward value (v) of a chosen bandit based on their prediction error (δ), i.e. the difference between the actual reward (r) and the expected reward for that trial:

$$v_{c_t,t+1} = v_{c_t,t} + \alpha\delta_t \quad \text{with} \quad \delta_t = r_t - v_{c_t,t}\ .$$

Herein, the indices t and $t+1$ denote the current and the next trial, respectively, and c_t the index of the bandit chosen on trial t. The parameter α denotes the learning rate, which was a free parameter in this model ranging between 0 and 1. The learning rate determines which fraction of the prediction error is used for updating. In contrast, the expected rewards of all unchosen bandits were not changed from one trial to the next, i.e. they remained constant until that bandit was chosen again. This trial-by-trial updating was initialized for each bandit with the same expected reward value v_1, which was another parameter of the model.

The second learning rule was the "Bayesian learner" model as described by Daw et al. (2006). This model implements the Kalman filter (Kalman, 1960; Kalman & Bucy, 1961; see also Anderson & Moore, 1979) as the Bayesian mean-tracking rule for the reward-generating diffusion process in the bandit task. First, this model assumes that subjects form an internal representation of the true underlying reward structure of the task. As described in section 2.4, the true reward structure followed a decaying Gaussian random walk determined by the parameters λ (decay parameter), ϑ (decay center), σ_o^2 (observation variance), and σ_d^2 (diffusion variance). In the cognitive model, subjects' estimations of

these parameters are denoted accordingly as $\hat{\lambda}$, $\hat{\vartheta}$, $\hat{\sigma}_o^2$, and $\hat{\sigma}_d^2$. Second, the model assumed that subjects update their reward expectations of the chosen bandit according to the Bayes rule (see 2.7.1). They start each trial with a prior belief about each bandit's mean payoff, which is normally distributed with mean $\hat{\mu}_{i,t}^{pre}$ and variance $\hat{\sigma}_{i,t}^{2\,pre}$ for bandit i on trial t. For the chosen bandit, this prior distribution is updated by the reward observation r_t, resulting in a posterior distribution with mean $\hat{\mu}_{i,t}^{post}$ and variance $\hat{\sigma}_{i,t}^{2\,post}$ according to:

$$\hat{\mu}_{c_t,t}^{post} = \hat{\mu}_{c_t,t}^{pre} + \kappa_t \delta_t \quad \text{with} \quad \delta_t = r_t - \hat{\mu}_{c_t,t}^{pre},$$

$$\hat{\sigma}_{c_t,t}^{2\,post} = (1 - \kappa_t)\hat{\sigma}_{c_t,t}^{2\,pre}.$$

Herein, the coefficient κ denotes the Kalman gain, which calculates for each trial t as:

$$\kappa_t = \hat{\sigma}_{c_t,t}^{2\,pre} / \left(\hat{\sigma}_{c_t,t}^{2\,pre} + \hat{\sigma}_o^2\right).$$

Similar to the learning rate parameter in the Delta rule, the Kalman gain determines the fraction of the prediction error that is used for updating. In contrast to the learning rate, however, the Kalman gain changes from trial to trial depending on the current variance of the prior expected reward distribution ($\hat{\sigma}_{c_t,t}^{2\,pre}$) and the estimated observation variance ($\hat{\sigma}_o^2$). The observation variance indicates how much the actual rewards vary around the (to be estimated) mean reward of a bandit and therefore reflects how reliable each trial's reward observation (each new data point) is for estimating the true underlying mean. If the prior variance is large compared to the estimated observation variance, i.e. if a subject's reward prediction is very uncertain while the reward observation is very reliable, the Kalman gain approaches 1 and a large fraction of the prediction error is used for updating. If, in contrast, the prior variance is very small compared to the estimated observation variance, i.e. if a subject's reward estimation is very reliable while reward observations are very noisy, then the Kalman gain approaches 0 and only a small fraction of the prediction error is used for updating. Similar to the Delta rule, the expected rewards (prior mean and variance) of all unchosen bandits are not updated *within* a trial, i.e. their posterior distributions equal their prior distributions for that trial. However, prior distributions of all four bandits are updated *between* trials based on the subject's belief about the underlying Gaussian random walk by:

$$\hat{\mu}_{i,t+1}^{pre} = \hat{\lambda}\hat{\mu}_{i,t}^{post} + \left(1 - \hat{\lambda}\right)\hat{\vartheta} \quad \text{and} \quad \hat{\sigma}_{i,t+1}^{2\,pre} = \hat{\lambda}^2\hat{\sigma}_{i,t}^{2\,post} + \hat{\sigma}_d^2\,,$$

wherein the indices t and $t+1$ denote the current and the next trial, respectively, and i the index of the bandit. The trial-by-trial updating process was initialized for all bandits with the same prior distribution $N(\hat{\mu}_1^{pre}, \hat{\sigma}_1^{2\,pre})$, whereby $\hat{\mu}_1^{pre}$ and $\hat{\sigma}_1^{pre}$ were two further parameters of the model.

Next, three different choice rules were used to model subjects' choices based on their expected rewards derived from either the Delta rule or the Bayesian learner rule. All choice rules were based on the commonly applied softmax function (McFadden, 1974; Sutton & Barto, 1998). The first model was the softmax function in its basic form without any bonus term (short: SM). According to this rule,

choices are probabilistically based on the relative expected reward values of all available choice options. The SM model has the form (a) if combined with the Delta rule and form (b) if combined with the Bayesian learner rule:

$$\text{(a)}\ P_{i,t} = \frac{exp(\beta v_{i,t})}{\sum_j exp(\beta v_{j,t})} \qquad \text{(b)}\ P_{i,t} = \frac{exp(\beta \hat{\mu}_{i,t}^{pre})}{\sum_j exp(\beta \hat{\mu}_{j,t}^{pre})}.$$

Herein, $P_{i,t}$ denotes the probability to choose bandit i on trial t, and Σ_j indicates a summation over all four bandits. The softmax β parameter, also called inverse temperature, reflects (inversely) the degree of randomness ("noisiness") in a subject's decisions: For small β, choices are very noisy, i.e. all actions have nearly the same probability irrespective of their relative expected values. The larger β gets, choices become less random and more and more value-driven (greedy). For extremely high β values, the choice probability for the option with the highest expected value approaches one, meaning choices are fully deterministic and always favoring this option (greedy strategy).

The second choice rule was a modified version of the softmax function called "softmax with exploration bonus" (short: SM+E), which was adopted from Daw et al. (2006). This model added an additional exploration bonus to the expected value of each bandit, which increased with the uncertainty of a bandit's outcome. Depending on the learning rule, different metrics were used to quantify that uncertainty. In the Bayesian learner rule, the uncertainty for a bandit i on trial t is directly quantified by the prior standard deviation of that bandit's expected reward distribution ($\hat{\sigma}_{i,t}^{pre}$). Since the Delta rule does not directly calculate these uncertainties, they were modeled based on a simple heuristic adopted from Speekenbrink and Konstantinidis (2015). According to that heuristic, a bandit's uncertainty increases linearly with the number of trials since it was last chosen. This is formalized as $(t - T_i)$, where T_i is the last trial before the current trial t in which bandit i was chosen. By incorporating these uncertainty metrics into the SM+E model, the model obtains form (a) if combined with the Delta rule and form (b) if combined with the Bayesian learner rule:

$$\text{(a)}\ P_{i,t} = \frac{exp(\beta[v_{i,t} + \varphi(t - T_i)])}{\sum_j exp(\beta[v_{j,t} + \varphi(t - T_j)])} \qquad \text{(b)}\ P_{i,t} = \frac{exp(\beta[\hat{\mu}_{i,t}^{pre} + \varphi\hat{\sigma}_{i,t}^{pre}])}{\sum_j exp(\beta[\hat{\mu}_{j,t}^{pre} + \varphi\hat{\sigma}_{j,t}^{pre}])}.$$

Herein, φ denotes the exploration bonus parameter, which reflects the degree to which choices are influenced by the uncertainty associated with each bandit. If φ is zero, choices are not influenced by these uncertainties and the SM+E model reduces to the simpler SM model. The larger φ gets, choices become more and more uncertainty-driven (assuming $\beta \neq 0$). Note that the softmax β is also often interpreted as an exploration parameter (e.g. Beeler et al., 2010, 2012; Beeler, 2012; Daw et al., 2006; Gershman, 2018; Humphries et al., 2012). However, since it simply reflects the noisiness of the decision in general, it describes only a form of random (undirected) exploration, while the φ parameter reflects a form of uncertainty-driven (directed or strategic) exploration (see Daw et al., 2006; Gershman, 2018).

The third choice rule was a novel extension of the SM+E model called "softmax with exploration bonus and perseveration bonus" (short: SM+EP). This version of the softmax rule included an extra perseveration bonus, which was a constant value (free parameter) only added to the expected value of the bandit chosen in the previous trial, but not to all other bandits. The SM+EP model has the form (a) if combined with the Delta rule and (b) if combined with the Bayesian learner rule:

(a) $P_{i,t} = \frac{exp(\beta[v_{i,t} + \varphi(t-T_i) + I_{c_{t-1=i}}\rho])}{\sum_j exp(\beta[v_{i,t} + \varphi(t-T_i) + I_{c_{t-1=j}}\rho])}$ (b) $P_{i,t} = \frac{exp(\beta[\hat{\mu}_{i,t}^{pre} + \varphi\hat{\sigma}_{i,t}^{pre} + I_{c_{t-1=i}}\rho])}{\sum_j exp(\beta[\hat{\mu}_{j,t}^{pre} + \varphi\hat{\sigma}_{j,t}^{pre} + I_{c_{t-1=j}}\rho])}$.

Herein, ρ denotes the perseveration bonus parameter and I an indicator function that equals 1 for the bandit that was chosen in the previous trial (indexed by c_{t-1}) and 0 for all other bandits.

Note at this point that the formalizations of the SM+E and SM+EP model included a bracket around the sum of expected rewards and bonus terms. Leaving out the bracket around this sum would mathematically result in the same model, but then the bonus parameters φ and ρ would obtain different values and interpretations (i.e. $\beta\varphi$ and $\beta\rho$, respectively). The formalization used here, which nested the bonuses within the softmax scheme (as described by Daw et al., 2006), has the advantage that bonuses can be directly interpreted in terms of reward value units in order to better compare all choice-influencing factors quantitatively with each other.

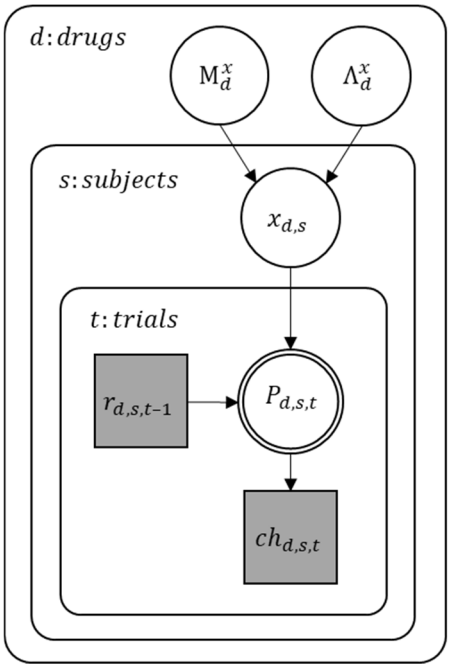
Taken together, by combing each learning rule with each choice rule, the following six cognitive models resulted for model comparison: Delta-SM, Delta-SM+E, Delta-SM+EP, Bayes-SM, Bayes-SM+E, and Bayes-SM+EP. The free and fixed parameters for each model are summarized in Table 1.

Parameters were estimated for each subject and drug condition using hierarchical Bayesian modeling. A graphical description of the modeling scheme is presented in Figure 5. The hierarchical modeling approach was chosen to improve the estimation of each subject-level parameter by assuming that these parameters are drawn from a group distribution (see 2.7.1; Gelman & Hill, 2007; Katahira, 2016; Kruschke & Vanpaemel, 2015). Parameters for the group distribution (mean and standard deviation) were estimated separately for each drug condition, which allowed the comparison of subject-level as well as group-level parameters between drugs. For all Bayesian learner models, the six parameters specifying subjects' estimation of the Gaussian random walk ($\hat{\lambda}, \hat{\vartheta}, \hat{\sigma}_o^2, \hat{\sigma}_d^2, \hat{\mu}_1^{pre}, \hat{\sigma}_1^{pre}$) – from here one referred to as "random walk parameters" – were initially fixed to constrain the free parameter space for these models, which largely facilitated estimation of the remaining choice parameters. Also, some of the Bayesian learner models did actually not converge with free random walk parameters, thus fixing these parameters was necessary in order to include all six cognitive models in the model comparison. In detail, the parameters $\hat{\lambda}$, $\hat{\vartheta}$, $\hat{\sigma}_o^2$, and $\hat{\sigma}_d^2$ were fixed to the values of the true underlying random walk parameters (see 2.4). The parameters $\hat{\mu}_1^{pre}$ and $\hat{\sigma}_1^{pre}$, specifying the mean and standard deviation of subjects' prior reward expectation for each bandit in the first trial, were fixed to $\hat{\mu}_1^{pre} = 50$ and $\hat{\sigma}_1^{pre} = 4$. Similarly, the parameter v_1 of the Delta rule models, specifying the expected reward value

Table 1. Free and fixed parameters of all six cognitive models.

	Delta rule		Bayesian learner rule	
SM	α, β	*fixed*: v_1	β	*fixed*: $\hat{\lambda}, \hat{\vartheta}, \hat{\sigma}_o^2, \hat{\sigma}_d^2, \hat{\mu}_1^{pre}, \hat{\sigma}_1^{pre}$
SM+E	α, β, φ	*fixed*: v_1	β, φ	*fixed*: $\hat{\lambda}, \hat{\vartheta}, \hat{\sigma}_o^2, \hat{\sigma}_d^2, \hat{\mu}_1^{pre}, \hat{\sigma}_1^{pre}$
SM+EB	$\alpha, \beta, \varphi, \rho$	*fixed*: v_1	β, φ, ρ	*fixed*: $\hat{\lambda}, \hat{\vartheta}, \hat{\sigma}_o^2, \hat{\sigma}_d^2, \hat{\mu}_1^{pre}, \hat{\sigma}_1^{pre}$

Note. Free parameters are only listed for the subject level. Note that hierarchical models contained for each free subject-level parameter x two additional free parameters (M^x, Λ^x) on the group level (see Figure 5). SM: softmax; SM+E: softmax with exploration bonus; SM+EP: softmax with exploration bonus and perseveration bonus; α: learning rate; β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter; v_1: initial expected reward value for all bandits; $\hat{\lambda}$: decay parameter; $\hat{\vartheta}$: decay center; $\hat{\sigma}_o^2$: observation variance; $\hat{\sigma}_d^2$: diffusion variance; $\hat{\mu}_1^{pre}$: initial prior mean of the expected reward for all bandits; $\hat{\sigma}_1^{pre}$: initial prior standard deviation of the expected reward for all bandits.



$$\mathrm{M}_d^x \sim uniform(x_{min}, x_{max})$$

$$\Lambda_d^x \sim half\text{-}Cauchy(0,1)$$

$$x_{d,s} \sim normal(\mathrm{M}_d^x, (\Lambda_d^x)^2)$$

replace x with model-specific parameters (see Table 1):

β with $\beta_{min} = 0, \beta_{max} = 3$

α with $\alpha_{min} = 0, \alpha_{max} = 1$

φ with $\varphi_{min} = -\infty, \varphi_{max} = \infty$

ρ with $\rho_{min} = -\infty, \rho_{max} = \infty$

Figure 5. Graphical description of the hierarchical Bayesian modeling scheme. In this graphical scheme, nodes represent variables of interest (squares: discrete variables; circles: continuous variables) and arrows indicate dependencies between these variables. Shaded nodes represent observed variables, here rewards (r) and choices (ch) for each trial (t), subject (s), and drug condition (d). For each subject and drug condition, the observed rewards until trial t-1 determine (deterministically) choice probabilities (P) on trial t, which in turn determine (stochastically) the choice on that trial. The exact dependencies between previous rewards and choice probabilities are specified by the different cognitive models and their model parameters (x). Note that the double-bordered node indicates that the choice probability is fully determined by its parent nodes, i.e. the reward history and the model parameters. As the model parameters differ between all applied cognitive models, they are indicated here by an x as a placeholder for one or more model parameter(s). Still, the general modeling scheme was the same for all models: Model parameters were estimated for each subject and drug condition and were assumed to be drawn from a group-level normal distribution with mean M^x and standard deviation Λ^x for any parameter x. Note that group-level parameters were estimated separately for each drug condition. Each group-level mean (M^x) was assigned a non-informative (uniform) prior between the limits x_{min} and x_{max} as listed above. Each group-level standard deviation (Λ^x) was assigned a half-Cauchy distributed prior with a location parameter 0 and scale 1. Subject-level parameters included α, β, φ, and ρ, depending on the cognitive model (see Table 1).

of each bandit in the first trial, was fixed to 50. Note that while these fix values for the initialization parameters were chosen somewhat arbitrarily, they only influence modeled choice behavior on the first few trials and thus have low impact on the overall model fit (see Daw et al., 2006). At this point, the reader is also referred to the discussion in section 6.5.2, in which the subject of fixing these random walk parameters to facilitate model fitting is reconsidered.

Bayesian modeling was performed using the software Stan (version 2.17.0; Stan Development Team, 2017b; see also Carpenter et al., 2017), operating from within the general statistical package R (version 3.4.3; R Core Team, 2017) with the interface rstan (version 2.17.2; Stan Development Team, 2017a). In Stan, posterior distributions of model parameters are stochastically approximated using Hamiltonian Monte Carlo sampling (Girolami & Calderhead, 2011). Sampling in Stan was performed with four chains, each chain running for 1000 iterations without thinning after a warmup period of 1000 iterations. Priors for all subject-level parameters were normally distributed with a parameter-specific mean (denoted by M^{x} for any parameter x) and standard deviation (denoted by Λ^{x} for any parameter x). The prior for each group-level mean was uniformly distributed within the limits as given in Figure 5. For each group-level standard deviation, a half-Cauchy distribution with location parameter 0 and scale parameter 1 was used as a weakly informative prior (see Gelman, 2006). For this specific half-Cauchy distribution, the 90th percentile is 6.31 and the 99th percentile is 63.66, thereby covering the most plausible values for the group-level standard deviations of each parameter, while also allowing for more extreme values.

Following parameter estimation, the six cognitive models were compared in terms of predictive accuracy using a Bayesian leave-one-out (LOO) cross-validation approach (Vehtari, Gelman, & Gabry, 2017). The LOO cross-validation approach measures pointwise out-of-sample predictive accuracy by repeatedly taking one data set ("testing set") out of the sample, refitting the model to the reduced data ("training set"), and then measuring how accurately the refitted model predicts the data of the testing set. This procedure is repeated as many times as there are data sets in the sample, with every single data set being used once as the testing set. Note that there are different ways here to define the scope of a testing set, e.g. as a single subject or a single trial. For the LOO analysis of the main study, a testing set was defined as the data of one subject under one drug condition, compounded over all trials. However, an alternative analysis was conducted in pilot study 2 (see 4.1.3), in which each cognitive model was fitted to the first 240 trials of each subject (training set) to predict that subject's choices in the last 60 trials (testing set). Since both approaches yielded largely similar results (see 4.2), only the first approach (LOO over subjects) was adopted in the main study.

Model comparison for the main study was performed using the data sets from all 31 subjects who completed the experiment, once combined over all drug conditions (yielding 93 data sets) and once separately for each drug condition (each with 31 data sets). In order to reduce computational burden, the R package loo (Vehtari et al., 2017) was used, which applies Pareto-smoothed importance sampling

to calculate LOO estimates that closely approximate exact LOO measures without refitting the model several times. Note that in pilot study 2, these LOO estimates were directly compared to the exact LOO measures, confirming that both approaches yielded very similar results (see 4.2). Applying this package, LOO estimates were calculated for each model fit based on its Stan output, using the log likelihood function evaluated at the sampled posterior parameter values. The log likelihood for each subject was calculated as the logarithmized product of choice probabilities (P) of the chosen bandits (indexed by c_t) compounded over trials t:

$$log\left(\prod_t P_{c_t,t}\right).$$

The detailed procedure of estimating LOO measures based on the subject-specific log likelihoods can be found in Vehtari et al. (2017). Since cross-validation measures like LOO are not biased in favor of more complex models (like ordinary goodness-of-fit measures), no penalty term is needed here to compensate for model complexity in order to prevent over-fitting.

Based on the results of the model comparison, the cognitive model with the highest predictive accuracy (Bayes-SM+EP) was then selected for further data analysis. Before that, however, the model was refit to the data with a different set of fixed values for the random walk parameters of the Bayesian learning rule. More specifically, while these parameters were initially fixed to their true (or arbitrarily chosen) values only for the model comparison (since some models did not converge with free random walk parameters, see above), these values were exchanged with the best-fitting parameter values obtained from pilot study 2 (except for σ_o^2, which was still fixed to its true value due to model degeneracy; see Daw et al., 2006). Therefore, each of these parameters was estimated once over all 16 subjects of the pilot data set (see 4.1.3), yielding the following posterior medians: $\hat{\lambda} = 0.93$, $\hat{\vartheta} = 46.0$, $\hat{\sigma}_d^2 = 6.6^2$, $\hat{\mu}_1^{pre} = 82.7$, and $\hat{\sigma}_1^{pre} = 3.6$ (see also Table A1 in the appendix). These posterior medians were then used as new fix values to refit the Bayes-SM+EP model to the data of the main study. As this new model fit showed even higher predictive accuracy than the previous one, it was chosen for all further model-based analyses.

On the one hand, parameter estimates of this model fit were used to test for DA drug effects on the behavioral level. First, posterior distributions of all six group-level parameters of this model (i.e. $\mathrm{M}^{\beta}, \mathrm{M}^{\varphi}, \mathrm{M}^{\rho}$ and $\Lambda^{\beta}, \Lambda^{\varphi}, \Lambda^{\rho}$) were compared between the DA drug conditions by calculating for each drug pair the difference between the drug-specific posterior distributions and then analyzing for each of these posterior differences the percentage of samples greater than zero. Additionally, it was analyzed if the 90% highest density interval (HDI) of these posterior drug differences included zero (see Kruschke, 2013, 2015). The 90% HDI defines that interval of a posterior distribution, in which every parameter value has a higher probability density than any value outside of the HDI and which contains 90% of its total mass. Hence, this interval reflects that part of the posterior distribution which contains the most credible drug difference values. Note that 90% HDIs are reported, since these are computational more stable than 95% HDIs, for which each end only relies on 2.5% of the posterior

samples (Gabry & Goodrich, 2018; Robert & Casella, 2005, p. 93). Second, DA drug effects on the subject-level posterior medians of each choice parameter (i.e. β, φ, ρ) were analyzed using a repeated measures ANOVA with the factor drug, followed by paired t-tests for each drug pair.

On the other hand, subject-level posterior medians of the three choice parameters (β, φ, ρ) were also used for further behavioral analyses, including the inverted-U analysis (see 2.9.2) and a correlation analysis with the percentage of exploratory trials (see below). In addition, these subject-level posterior medians were used to generate trial-by-trial regressor for the model-based fMRI analysis (see 2.8.3) as described by Daw (2011), including the expected value ($\hat{\mu}^{pre}$) and uncertainty ($\hat{\sigma}^{pre}$) of the chosen bandit, the reward prediction error (δ), and the overall uncertainty (denoted here as $\Sigma\hat{\sigma}^{pre}$). The latter variable was computed by summing for each trial the uncertainty, i.e. the prior standard deviation ($\hat{\sigma}^{pre}$), over all four bandits. Moreover, some trial-by-trial variables – including the expected value ($\hat{\mu}^{pre}$), exploration bonus ($\varphi\hat{\sigma}^{pre}$), and perseveration bonus ($I\rho$) of each bandit – were also used to classify subjects' choices into different choice types based on two different classification schemes: First, a binary classification scheme adopted from Daw et al. (2006), which divides choices into either exploitations or explorations. According to this scheme, a choice is exploitative if the bandit with the highest expected value was chosen, and exploratory if one of the other bandits was chosen. Second, a trinary classification scheme, which divides choices into exploitations, directed explorations, and random explorations. According to this scheme, a choice is exploitative if the bandit with the highest expected value was chosen, or the bandit with the highest sum of expected value plus perseveration bonus. All the remaining (i.e. non-exploitative) trials are classified as either directed explorations, i.e. trials in which the bandit with the highest exploration bonus was chosen, or random explorations, i.e. trials in which not the bandit with the highest exploration bonus was chosen. The choice types derived from both classification schemes were also used as regressors in the model-based fMRI analysis (see 2.8.3). Note also that Pearson correlations between all model-based variables that were used as fMRI regressors are reported in Table A4 of the appendix.

Furthermore, the choice classification schemes were used to calculate the percentage of overall explorations (according to the binary classification) and the percentage of random and directed explorations (according to the trinary classification) for each subject over all trials per session in order to analyze pairwise Pearson correlations between these percentages and the subject-specific posterior medians of all choice parameters (β, φ, ρ). Note that data from the placebo condition (n=31) and pilot study 2 (n=16) were combined to increase the sample size for this correlation analysis (n=47). Finally, DA drug effects on the percentage of exploratory trials (overall, random, and directed) were analyzed by performing on each of these dependent variables a repeated measures ANOVA with the factor drug, followed by paired t-tests for each drug pair.

2.8 Functional magnetic resonance imaging (fMRI)

2.8.1 Introduction to functional magnetic resonance imaging

Since its development in the early 1990s, functional magnetic resonance imaging (fMRI) has gained immense popularity as a research tool to investigate human brain function (Poldrack, Mumford, & Nichols, 2011). FMRI is a noninvasive technique that measures brain activity from local changes in blood oxygenation via magnetic resonance (MR) imaging (see below; Heeger & Ress, 2002; Logothetis, 2002; Sprawls, 2000). It is based on the fact that neuronal activation leads to an increased blood flow through the active brain area, called hemodynamic response, whereby the additional blood carries more oxygen than is actually needed by the active neurons (Fox & Raichle, 1986; Fox, Raichle, Mintun, & Dence, 1988). This relative increase in the local blood oxygenation level changes the local magnetic properties of the tissue, which can be measured by fMRI as the so called "blood oxygenation level dependent" (BOLD) signal (Ogawa et al., 1992; Ogawa, Lee, Kay, & Tank, 1990; see also Heeger & Ress, 2002). More specifically, the BOLD signal relies on the different magnetic properties of deoxygenated and oxygenated hemoglobin in the blood: While deoxygenated hemoglobin is strongly paramagnetic and induces inhomogeneities in a local magnetic field, oxygenated hemoglobin is weakly diamagnetic and shows little effect on a magnetic field (Logothetis, 2002). The shift in the ratio of paramagnetic to diamagnetic hemoglobin in response to neuronal activity can be detected via MR imaging.

MR imaging is a technique that uses a magnetic field and radio frequency signals to visualize anatomical structures and physiological processes within the body (for details see Sprawls, 2000). In an MR scanner, a subject's head is exposed to a strong magnetic field (B_0), whereby field strengths of 1.5 or 3.0 Tesla are commonly applied. As the hydrogen nuclei (protons) in the tissue are constantly rotating around their own axes, called "spin", each of them exhibits a magnetic moment with random orientation. Application of the external magnetic field changes the protons' magnetic moments from a random orientation to an orientation in which they are aligned either parallel or anti-parallel to the external field. As the parallel alignment reflects the lower-energy state, more protons are aligned parallel than anti-parallel to B_0. Furthermore, the protons start to "precess" in the external magnetic field, meaning that their spin "wobbles" in a cone-shaped form around the axis of B_0. Overall, the result of this field-induced spin alignment and precession is a net magnetization parallel to B_0 known as longitudinal magnetization. In this state of longitudinal magnetization, radiofrequency (RF) pulses are transmitted to the tissue, which deflect the protons from their alignment along the B_0 axis. A fraction of the protons is flipped from the parallel to the anti-parallel state, which results in a reduction of the net longitudinal magnetization. In addition, the RF pulse leads to a synchronization of precessing protons, which results in a net magnetization orthogonal to the B_0 axes, called transverse magnetization. These RF pulses are repeated in short intervals, referred to as "repetition time" (TR) of the MR measurement. When the RF pulse is switched off, the protons start to relax, whereby relaxation occurs in two different ways. First, a fraction of the protons fall back into the lower energy state of parallel alignment to B_0, thereby restoring longitudinal magnetization, which is called longitudinal (or

T1) relaxation. Second, synchronized precessing protons start to desynchronize again, thereby reducing transverse magnetization, which is called transverse (or T2) relaxation. These relaxation processes produce electrical signals which can be detected by RF receiver coils placed around the subject's head, whereby these signals are localized in space by the use of magnetic field gradients. As T1 and T2 relaxation times vary between different tissues, they generate signal contrasts which can be visualized in different types of MR images (Currie, Hoggard, Craven, Hadjivassiliou, & Wilkinson, 2013; Sprawls, 2000). T1-weighted images primarily rely on differential T1 relaxation times between different tissues and are often used in fMRI to generate high-resolution structural brain images to aid anatomic localization of BOLD signals. In contrast, T2- and T2*-weighted images predominantly rely on differential transversal relaxation times. Thereby, transverse relaxation is further distinguished into T2 relaxation, which is caused by local spin-spin interactions in the tissue, and T2* relaxation, which is additionally caused by local field variations (inhomogeneities) within B_0 (Chavhan, Babyn, Thomas, Shroff, & Haacke, 2009). Importantly, the T2* relaxation time is also influenced by changes in blood oxygenation following neuronal activation, since the relative decrease in paramagnetic deoxyhemoglobin reduces local field inhomogeneities and prolongs the T2* relaxation time (Logothetis, 2002; Ogawa et al., 1990). Thus, the BOLD signal can be visualized in a T2*-weighted image, in which active brain areas temporarily exhibit a stronger signal and appear brighter in the image (Sprawls, 2000). For fMRI, T2*-weighted images are usually acquired with a very rapid MR imaging technique called echo planar imaging (EPI; Mansfield, 1977; Poustchi-Amin, Mirowitz, Brown, McKinstry, & Li, 2001).

To arrive from a raw fMRI scan to an interpretable image of brain activity, multiple analysis steps are required. First, several preprocessing steps are usually performed to reduce noise, improve data quality, and transform images to a common anatomical space for later analyses. In short, these steps include quality control, distortion correction, motion correction, slice timing correction, temporal filtering, spatial normalization, and spatial smoothing (for details see Poldrack et al., 2011; Soares et al., 2016; see also 2.8.3). The preprocessed images are then subjected to statistical analysis, which usually includes the steps of statistical modeling, inference, and visualization of results in statistical maps (for details see Friston, Ashburner, Kiebel, Nichols, & Penny, 2006; Poldrack et al., 2011). Statistical modeling of fMRI data commonly relies on the general linear model (GLM) approach (Friston et al., 1994; Kiebel & Holmes, 2006). The GLM approach basically performs an independent multiple regression analysis for every single voxel (i.e. volume element) in the fMRI scan. The model assumes thereby that the observed BOLD time series (Y) of a given voxel can be expressed as a linear combination (weighted sum) of one or more experimental design variables (X_1, X_2, …, X_n), each weighted by a regression coefficient (β_1, β_2, …, β_n), plus a constant (the intercept β_0) and a random error term (ε). In vector notation, this GLM can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon.$$

Herein, the different design variables (X_i), also called "regressors", are typically task-related. They may represent onset times of different stimulus events in the experiment (e.g. trial onsets, reward onsets) or different parametrical modulators of such events (e.g. reward values). As different regressors may be correlated with each other to some degree, they are often orthogonalized in the GLM (see Mumford, Poline, & Poldrack, 2015; Poldrack et al., 2011). Orthogonalization results in uncorrelated regressors (i.e. no shared variance), such that only the variability unique to one specific regressor determines the estimate of its regression coefficient. Before the regressors are entered into the GLM analysis, they are typically convoluted with a canonical hemodynamic response function (HRF), which describes the time course of a typical BOLD response to a single brief stimulus. The canonical HRF rises within 1-2 s and peaks around 5 s after stimulus onset, followed by a small undershoot (negative peak) that lasts up to 20-30 s after stimulation (see Henson & Friston, 2006; Poldrack et al., 2011, chapter 5).

After model specification, the regression coefficients ("betas") for each regressor are then estimated by least squares optimization, which minimizes the squared distance between the observed data and the values predicted by the GLM. The resulting beta estimates reflect how strongly each regressor contributed to the observed BOLD time series in a given voxel, controlling for all other regressors in the model. Beta estimates for all recorded voxels can be collectively visualized in the so called "beta image" of a regressor, and different beta images can be linearly combined to obtain so called "contrast (con) images". Next, these beta and con images can be used for statistical inference (hypothesis testing), which can be performed either on the subject level ("first-level analysis") or across subjects on the group level ("second-level analysis"; see Friston et al., 2006; Poldrack et al., 2011). On the second level, random effects models are often applied, which account for both within-subject and between-subject variability in the data (Friston, Stephan, Lund, Morcom, & Kiebel, 2005; Penny & Holmes, 2006). As the term "general" implies, the GLM approach can be used for various types of statistical tests, including paired and unpaired t-tests, analysis of variance (ANOVA), and analysis of covariance (ANCOVA). For example, by subtracting the beta images of two task-related regressors (e.g. explore minus exploit), it can be tested for each voxel if one task condition elicited a significantly stronger or weaker BOLD response than the other. The results of these tests are usually visualized in statistical parametric maps, which show the according test statistic (e.g. t- or F-values) for each voxel in the fMRI scan. For better visualization, these maps are often thresholded at a certain value, color-scaled, and overlaid onto a structural (T1) brain image of the subject or a mean T1 image of the group.

As hypothesis testing is performed simultaneously for every single voxel in an fMRI scan, which typically contains more than 100 000 voxels across the brain, this testing approach has a high chance of yielding an accordingly large number of false positive results (Type I errors), which is known as the "multiple testing problem" (Nichols, 2012). Different approaches have been developed to correct for this problem by using p-values adjusted for multiple comparisons instead of uncorrected p-values (see Brett, Penny, & Kiebel, 2006; Nichols, 2012; Nichols & Hayasaka, 2003). One such approach is called the "familywise error" (FWE) correction, whereby the familywise error rate denotes the probability to

obtain at least one false positive result in a family of statistical tests. Hence, using FWE-corrected p-values and a threshold of $p < .05$ in a whole-brain fMRI analysis ensures that the probability to obtain at least one false positive results across all tested voxels is less than 5 %. Alternatively, the FWE correction can also be applied to an a priori selected brain region of interest in order to reduce the number of tests to correct for, which is known as "small volume (FWE) correction" (Brett et al., 2006; Poldrack et al., 2011). In fMRI, the FWE approach is commonly based on Gaussian random field theory to account for the fact that statistical tests on single voxels in an fMRI scan are not actually independent of each other due to spatial correlations in the data (Nichols & Hayasaka, 2003; Worsley, 2006).

Finally, model-based fMRI combines the fMRI approach with cognitive modeling, yielding a powerful tool to investigate the neural correlates of cognitive processes underlying observed behavior. In this approach, trial-by-trial estimates for latent (unobservable) cognitive variables derived from the model, such as the subjective value or uncertainty associated with a given choice option, can be entered as parametric regressors into the GLM to investigate the neural correlates of these variables (Daw, 2011; Gläscher & O'Doherty, 2010). For instance, model-based fMRI has enabled to pinpoint the neural correlates of the reward prediction error signal from reinforcement learning models by showing that this signal positively correlates with the BOLD response in striatal brain regions (Dreher, 2013; O'Doherty et al., 2004; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003). Importantly, also the neural signatures of explore/exploit decisions have been analyzed by using this method, as the classification of choices into explorations and exploitations is mostly based on model-derived subjective value estimates (see 1.1.4; e.g. Addicott et al., 2014; Daw et al., 2006; Laureiro-Martínez et al., 2014). Note that the procedure for such a model-based fMRI analysis of explore/exploit decisions is described in more detail below (see 2.8.3).

For a comprehensive introduction to MR imaging and fMRI, the reader is referred to the textbooks of Sprawls (2000), Poldrack et al. (2011), and Friston et al. (2006). More details on model-based fMRI and its application in decision neuroscience can be found in the reviews by Gläscher and O'Doherty (2010), Daw (2011), and O'Doherty et al. (2007).

2.8.2 fMRI data acquisition

Functional imaging data were acquired on a Siemens Trio 3T scanner (Erlangen, Germany) equipped with a 32 channel head-coil. For each subject and drug condition, four blocks à 75 trials were recorded for the bandit task. The first five scans of each block served as dummy scans to allow for magnetic field saturation and were discarded. Functional volumes were recorded using a T2*-weighted EPI sequence. Each volume consisted of 40 slices with 2 mm isotropic voxels and 1 mm gap, acquired with a repetition time (TR) of 2470 ms, an echo time (TE) of 26 ms, and a flip angle of 80°. In addition, a high-resolution structural image was acquired for each subject at the end of the first fMRI session, using a T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence with 1 mm isotropic voxels and 240

slices. The experimental task was projected onto a mirror attached to the head coil and participants responded by using a button box with four buttons held in the right hand (see 2.4).

2.8.3 fMRI data analysis

Preprocessing and statistical analysis of fMRI data was performed using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK). The preprocessing included four steps. First, to correct for subject motion, functional images of each subject were realigned to the first functional image of the placebo condition using a six-parameter affine transformation and unwarped to correct for the interaction between motion and distortion. Second, functional images were slice time corrected to the onset of the middle slice to correct for the shifted acquisition time of the slices within a volume. Third, all images were spatially normalized to Montreal Neurological Institute (MNI) space using the DARTEL toolbox (Ashburner, 2007). Therefore, the structural T1 image of each participant was first coregistered to the mean functional image (generated during realignment) and segmented into grey matter, white matter, and CSF. The resulting segmented images were then used within the DARTEL toolbox to normalize the structural T1 image and functional images of each subject to MNI space, and to resample functional images to 1.5 mm isotropic voxels. The normalized T1 images were subsequently averaged across all subjects to be used as a mean structural scan for visualization purposes (see below). Fourth, functional images were spatially smoothed using a Gaussian kernel of 6 mm full-width at half-maximum (FWHM).

For the first-level analysis of fMRI data, a general linear model (GLM) was created for each subject and drug condition, concatenated over all four blocks of the bandit task within one drug condition. For each trial in which a bandit was chosen, two different time points were included in the model: the time of the trial onset and the time of the outcome presentation. GLM regressors for these time points were created by convolving the event onsets, modeled by a stick function of zero duration, with the canonical hemodynamic response function (HRF) as implemented in SPM12. In addition, several parameteric modulators of these onset regressors were included in the model, which were also convolved with the HRF. First, the type of each choice (1 = *explore*, 0 = *exploit*) was entered as a parametric modulator of the trial onset regressor. Choice types were derived from the cognitive model according to the binary classification scheme used by Daw et al. (2006), as described above (see 2.7.2). Second, the reward prediction error δ, also derived from the cognitive model (see 2.7.2), was entered as a first parameteric modulator of the outcome onset regressor, and the outcome value (the number of points gained on each trial) as its second parametric modulator. For trials in which no bandit was chosen, the model contained an additional error regressor constructed by convolving the onsets of these trials with the HRF. Furthermore, four sessions constants (not convolved with the HRF) were included in the model as regressors for the four concatenated task blocks. Low-frequency noise was removed by employing a temporal high-pass filter with a cut-off frequency of 1/128 Hz, and a first-order autoregressive model AR(1) was used to remove serial correlations. After each first-level GLM

was estimated, regressor-specific contrast images were created for each subject and drug condition for the following five regressors of interest: trial onset, outcome onset, choice type, prediction error, and outcome value.

Next, the first-level results were taken to a second-level random effects analysis to allow for statistical inference on the group level. For each regressor of interest, the subject- and drug-specific contrast images were submitted to a flexible factorial model in SPM12, including the factors drug (3 levels, within-subject), subject (31 levels), and a constant. The drug factor was specified as containing dependent observations of equal variance, and the subject factor as containing independent observations of equal variance. For each contrast-specific second-level analysis, a t-contrast image was created that tested for the main effect of that specific contrast over all subjects and drug conditions, calculated by weighting each drug level by one and each subject level by 3/31 (see Gläscher & Gitelman, 2008). Note that for the choice type regressor, computing this t-contrast with positive weights only tests for the direction *explore > exploit*, as explore trials were coded with 1 and exploit trials with 0 in this regressor. Hence, the same t-contrast was also computed with negative weights in order to create a t-contrast image for the comparison *exploit > explore*.

In addition to the main (first) GLM, two alternative GLMs were created and analyzed accordingly on the first and second level. Both alternative GLMs only differed from the main GLM with respect to the regressors modeled at trial onset, while the remaining regressors remained the same. Whereas the main GLM included one trial onset regressor with one parametrical modulator (*explore*/*exploit*), the second GLM included instead three trial onset regressors: one for directed explorations (*directed*), one for random explorations (*random*), and one for exploitations (*exploit*). These three choice types were defined according to the trinary classification scheme as described above (see 2.7.2). For this GLM, the second-level random effects analysis included the t-contrasts *directed > exploit*, *random > exploit*, *directed > random,* and *random > directed.* The third GLM included one trial onset regressor with two parametric modulators: the *expected value* ($\hat{\mu}^{pre}$) and *uncertainty* ($\hat{\sigma}^{pre}$) of the chosen option (in that order), both derived from the cognitive model as described above (see 2.7.2). Note that also a second version of this GLM was analyzed with the reverse order of both parametric modulators, which yielded substantially the same results as the first version and is thus not further considered here. For the third GLM, t-contrasts for both parameteric modulators, i.e. *expected value* and *uncertainty*, were included in the second-level random effects analysis.

To test for DA drug effects across subjects, an F-contrast image was created for each contrast-specific second-level analysis (see above) with the weights [1 -1 0; 0 1 -1] over the three drug levels [P D H] and zero weights for all 31 subject levels. Note that this F-contrast tests for the main effect of the drug condition analogous to a repeated measures ANOVA (see Henson & Penny, 2005). Since this F-test performs an undirected (two-sided) comparison, F-contrasts for the main and second GLM are denoted accordingly (i.e. without direction) as *explore vs. exploit* (main GLM), *directed vs. exploit*, *random vs. exploit*, and *directed vs. random* (second GLM). In addition to the second-level ANOVA,

a second-level regression analysis was conducted for each drug pair to test whether DA drug effects on exploration-specific brain activity were linearly predicted by DA drug effects on exploratory behavior. For this, the subject- and drug-specific contrast images for *explore vs. exploit* were used to calculate for each subject the difference image of this contrast for a given drug pair (P-D, P-H, or D-H). These difference images were then used in the second-level regression analysis, including the subject-specific drug differences of the exploration bonus parameter (φ posterior medians; see 2.7.2) for the same drug pair as explanatory variable. The same kind of regression analysis was also performed for the contrasts *directed vs. exploit* and *random vs. exploit* of the second GLM.

Finally, a fourth GLM was created for an additional post-hoc exploratory analysis. This fourth GLM differed from the main GLM only with respect to the parametric modulator of the trial onset regressor, replacing the binary variable choice type (*explore*/*exploit*) by a continuous model-based variable termed *overall uncertainty* ($\Sigma\hat{\sigma}^{pre}$), which is the summed uncertainty ($\hat{\sigma}^{pre}$) over all four bandits (see 2.7.2). The contrast images for the *overall uncertainty* regressor were then used in a second-level random effects analysis (as described above) to test for drug differences in the parametric effects of this regressor across subjects. Since this post-hoc analysis specifically focused on a comparison of the placebo and L-dopa condition (based on the behavioral findings, see 5.2), the first second-level model only included these two drug conditions. Based on this model, different t-contrast images were created to test for the parameteric effects of this regressor in the placebo condition alone, and for its differential parametric effects between both drug conditions (*placebo* > *L-dopa*, *L-dopa* > *placebo*). For completeness, also a second-level analysis with all three drug conditions was performed to test for the remaining pairwise drug effects accordingly (*placebo* > *haloperidol*, *haloperidol* > *placebo*, *L-dopa* > *haloperidol*, *haloperidol* > *placebo*). Finally, also a second-level regression analysis (as described above) was performed for this regressor.

All fMRI results are reported at a threshold of $p < .05$, FWE-corrected for the whole brain volume, unless stated otherwise. In addition, results of the second-level ANOVA and regression analysis for the first and second GLM (i.e. exploration-specific contrasts) were also analyzed using small volume FWE correction ($p < .05$) for seven regions that have previously been associated with exploratory choices: the left/right FPC and left/right IPS (Daw et al., 2006), as well as the dACC and left/right AI (Blanchard & Gershman, 2018). Regions used for small volume correction were defined by a 10 mm radius sphere around the respective peak voxel reported by the previous studies (see Table A5 in the appendix). For display purposes, an uncorrected threshold of $p < .001$ was used (unless stated otherwise), and activation maps were overlaid on the mean structural scan of all participants.

2.9 Further behavioral data analysis

2.9.1 Model-free choice variables

In addition to the cognitive modeling analysis, also four model-free measures of choice behavior were analyzed. First, the total monetary payout of the task (denoted here as *payout*) was used as a model-free performance measure, assuming that a successful balance between exploitation and exploration results in a high overall payout. The second model-free variable was the percentage of choices in which the bandit with the highest actual payoff was chosen (denoted here as *% best bandit*). A high value of *% best bandit* over time is also assumed to indicate a successful explore/exploit trade-off, in which a subject exploits a bandit as long as it pays out best, but also knows when to switch to the next bandit that pays out best through occasional exploration. The third model-free variable was the mean rank of all choices (denoted here as *mean rank*), whereby rank refers to the actual payoff of the bandit from lowest (1) to highest (4). This measure is similar to *% best bandit*, but considers not only the fraction of choices in which the best bandit was chosen, but also where the second or third best bandits were chosen over the fourth bandit. For example, even if two subjects chose the first best bandit equally often, their average choice ranks might still differ if one subject selected the second and third best bandit more often than the other subject. If choices are totally random, the expected value for *mean rank* is 2.5, whereas it should be close to 4.0 for an optimal performer. The fourth model-free variable was the percentage of all choices in which the bandit was switched (denoted here as *% switches*). While switching to another bandit is not per se an indicator of an exploratory choice, it may still be assumed to highly correlate with the percentage of exploratory choices. Note, however, that a switch can also be an exploitative choice, e.g. if a subject switches back to the bandit with the highest expected reward after having explored other options. Note further that while the first three model-free variables (*payout*, *% best bandit, mean rank*) were all assumed to be indicators of a successful explore/exploit balance in terms of maximizing the overall payout, the fourth variable (*% switches*) is assumed to indicate the extent of exploration alone rather than the balance between exploration and exploitation. For each of these model-free choice variables, the effect of the drug condition was tested using a repeated measures ANOVA with the factor drug.

2.9.2 Inverted-U analysis

The inverted-U analysis containted two steps: In a first step, it was tested whether individual differences in explore/exploit behavior under drug-free conditions were predicted by the individual DA baseline according to an inverse quadratic (inverted-U-shaped) relationship. In a second step, it was tested whether DA drug effects on explore/exploit behavior were modulated by the individual DA baseline, as also predictied by the inverted-U hypothesis. For both steps, the sEBR and WMC were used as behavioral proxy measures for the individual DA baseline. All parts of this analysis were performed in R (version 3.4.3; R Core Team, 2017).

For the first step, data from the placebo condition (n=31) and pilot study 2 (n=16) were combined to increase the sample size to 47 subjects. Two regression models were then fitted to these data. First, a regression model termed "linear model" (LM), which only tested for a linear relationship between a given behavioral choice measure and DA proxy measure according to:

$$behavioral\ choice\ measure\ =\ \beta_0 +\ \beta_1\ DAproxy + \varepsilon,$$

with ε denoting a random error term. Second, a regression model termed "quadratic model" (QM), which also tested for a quadratic relationship between these variables according to:

$$behavioral\ choice\ measure\ =\ \beta_0 +\ \beta_1\ DAproxy\ +\ \beta_2\ DAproxy^2 + \varepsilon.$$

As behavioral choice measures, different model-based and model-free variables were used. Model-based measures of choice behavior included the subject-level posterior medians for all choice parameters of the Bayes-SM+EP model, i.e. the softmax parameter (β), the exploration bonus parameter (φ), and the perseveration bonus parameter (ρ). Note that while the ρ parameter does not reflect explore/exploit behavior, it was nonetheless included in the analysis for exploratory purposes. Model-free measures of choice behavior included the four variables introduced in section 2.9.1, i.e. *payout*, *% best bandit, mean rank,* and *% switches*. The first DA proxy used for this regression analysis was the sEBR (see 2.3.1). For the second DA proxy, a principal component analysis (PCA) was performed over the z-transformed scores of the following WMC tasks: the Rotation Span Tak (absolute scores), the Operation Span Task (absolute scores), and the Listening Span Task (span scores). The Digit Span Backward Task was not included in this PCA because of too many missing values (zero scores), since 9 out of 47 subjects misunderstood the task instructions and performed the forward instead of the backward version of the task. The PCA was performed using the R function *prcomp*, and the first principal component of this PCA, denoted WMC_{PCA}, was used as DA proxy in the regression analysis. For exploratory purposes, each of the three WMC task scores included in this PCA was also used separately as a DA proxy in the regression analysis.

The regression coefficients β_0, β_1, and β_2 of all models were estimated using the R function *lm*, which fits linear models based on the ordinary least squares method. After fitting all regression models to the data, a LOO cross-validation approach (see 2.7.2) was used to compare the LM fit and QM fit for each pair of variables. The LOO model comparison was performed as follows: For each subject in the sample (n=47), both regression models were fitted to a reduced data set excluding that subject (training set). Then, predictive accuracies of both models were calculated for the left-out subject (testing set), using the squared distance between the true value for that subject and the predicted value based on the respective model fit. Finally, the squared distances for all subjects were averaged to yield the overall LOO measure of each model, which were then compared between the LM and QM. Additionally, the resulting p-values for the β_2 coefficients, which test for the null hypothesis $\beta_2 = 0$, were examined for each QM.

The second step of this inverted-U analysis was based on the data of all subjects who completed all three drug conditions of the main study (n=31). For this analysis, only the three model-based choice parameters (β, φ, ρ) were used as dependent variables to keep the number of statistical tests in a reasonable range (see below). DA proxies were the same as in the first step, i.e. the sEBR, WMC_{PCA}, and the three separate WMC task scores (see above). For each behavioral choice measure, the magnitude of the DA drug effect was first calculated for each subject by computing the difference of this measure for each drug pair, i.e. placebo minus L-dopa (P-D), placebo minus haloperidol (P-H), and L-dopa minus haloperidol (D-H). These drug differences were then plotted against each DA proxy. The turning point of the inverted-U curve was approximated by the sample's median value and marked in the plots by a vertical line. Based on the assumptions of the inverted-U hypothesis, the direction and magnitude of the DA drug effects were expected to differ between subjects below and above the turning point of the inverted-U curve (see 1.2.4; Figure 3), which was visually examined in the plots. Furthermore, for each DA proxy measure, the subject sample was split at its median value into a low (n=16) and a high (n=15) DA baseline group, and DA drug effects (P-D, P-H, and D-H) on each behavioral choice measure were compared between both groups using two-sample t-tests.

2.9.3 Control variables

Several control variables were tested for DA drug effects. The first set of control variables was measured during the post-fMRI testing (see 2.5) and comprised 18 variables in total: the sEBR, the total scores of the Digit Span Forward and Backward, and 15 attentional performance measures from the TAP, including reaction time medians, reaction time standard deviations, and error rates for each of the subtests Alertness, Go/NoGo, and Flexibility, as well as the index of phasic arousal for the subtest Alertness. Each of these 18 control variables was used as a dependent variable in a univariate repeated measures ANOVA with the factor drug and data from 31 subjects. Note that due to missing data, the ANOVA for the sEBR and all variables of the subtest Flexibility included only 30 subjects, and for the Digit Span Backward only 29 subjects. Since none of the 18 control variables showed a significant ($p < .05$) drug main effect in this ANOVA, no further t-tests were performed.

The second set of control variables, measured at different time points throughout each fMRI session, comprised a total of 19 variables on subjective mood and physical wellbeing (see 2.6). Mood parameters included ratings on the subscales alertness, contentedness, and calmness from the VAS (Bond & Lader, 1974) and ratings on the dimensions pleasure, arousal, and dominance as assessed by the SAM (Lang, 1980). Physical wellbeing parameters included pulse, blood pressure (systole and diastole), and ratings on ten potential drug side effects (see 2.6). For statistical testing, ratings on the ten potential side effects were summed to yield a "side effects sum score" for each time point. Mood variables were obtained at three different time points: before ingestion of the first pill (t_0), directly before testing in the fMRI scanner (t_1), and directly after testing in the fMRI scanner (t_2). Physical wellbeing variables were assessed at four different time points: before ingestion of the first pill (t_0),

one hour after ingestion of the first pill (t_1), directly before testing in the fMRI scanner (t_2), and at the very end of the post-fMRI testing (t_3). To test for drug effects on these variables, their scores at later time points (t_1, t_2, t_3) were first subtracted by the baseline score (t_0) for each drug condition. Each of these difference scores (t_1-t_0, t_2-t_0, t_3-t_0) was then used as dependent variable in a univariate repeated measures ANOVA with the factor drug. For each control variable that showed a significant drug effect ($p < .05$) in this ANOVA, three paired t-tests were conducted to test for significant mean differences between each pair of drug conditions, i.e. placebo vs. L-dopa, placebo vs. haloperidol, and L-dopa vs. haloperidol.

For completeness, also reaction times in the bandit task were tested for DA drug effects. For each subject, the mean and median reaction time was calculated across all 300 trials per drug condition, and both measures were used as a dependent variable in a univariate repeated measures ANOVA with the factor drug.

2.9.4 Drug guesses

Drug guesses from each subject at the end of each fMRI session were analzyed in order to test if participants were able to guess the drug they received above chance level. Since subjects were instructed to make their drug guesses on each fMRI session independently of the other sessions (see 2.6), all drug guesses were assumed to be independently of each other for the analysis. First, subjects' drug guesses were classified as correct or incorrect and it was tested whether the proportion of correct guesses over all subjects and drug sessions exceeded the one expected by chance alone (which was 1/3, since there were three guessing options of which one was correct). Second, the number of correct guesses per subject was counted (0, 1, 2, or 3) and it was tested whether its frequency distribution over all subjects differed from the one expected for random guessing using a chi-squared test (with Monte Carlo approximation). Third, the frequency of each drug guess ("placebo", "L-dopa", or "haloperidol") over all subjects was counted separately for each drug condition and it was tested whether these frequencies differed between the three drug conditions, also using a chi-squared test. Finally, it was analyzed if subjects' confidence ratings for drug guesses differed between the three drug conditions using a repeated measures ANOVA with the factor drug, or if they differed between correct and incorrect guesses using a two-sample t-test. Note that only 29 subjects were included in the repeated measures ANOVA due to missing confidence ratings for two subjects. For the same reason, the two-sample t-test only included confidence ratings for 30 correct vs. 61 incorrect guesses.

3 Pilot study 1

The aim of this pilot study was to assess the temporal stability (retest reliability) and interindividual variability of different potential proxy measures for DA, following the rationale that a reliable predictor of baseline DA function should be relatively stable over time and variable between subjects. Tested DA proxies included the spontaneous eye blink rate (sEBR) and various measures of working memory capacity (WMC). Based on the results of this pilot study, some of the tested measures were then selected to be used as DA proxies in the main study.

3.1 Study-specific methods

3.1.1 Participants

In total, 16 healthy male subjects participated in the study (aged 19 to 32 years, $M = 24.31$, $SD = 3.42$), of which 15 completed both experimental sessions. One subject dropped out after the first session and was not included in the analysis. Furthermore, one subject scored zero in one of the tasks (Digit Span Backward) due to a misunderstanding of the task instruction and was excluded from the data analysis of the Digit Span Backward task. Participants were recruited via an online job portal (www.stellenwerk.de) and were screened for the following inclusion criteria: male, age 18-35 years, right-handed, fluent German in speaking and writing, normal or corrected to normal vision, no hearing impairments, no major past or present psychological, neurological, or physical disorders, no regular consumption of prescription drugs. Before participating in the study, all subjects provided informed written consent, and study procedures were approved by the local ethics committee (Hamburg Medical Council). Participants were paid after the experiment with 10 € per hour, which resulted in an average total payout of about 60 € per subject.

3.1.2 General procedure

To study the retest reliability of the sEBR and different WMC tasks, each subject performed a testing session (see 3.1.3) similar to the baseline screening of the main study on two separate days. The two sessions of each subject were scheduled exactly one week apart and at the same time of day. Sessions started between 10 am and 4 pm and lasted about 2.5 hours on average, separated by small breaks. Sessions were conducted in groups of four in the same experimental room, each subject sitting in front of a computer screen, separated by partition walls. An experimenter was present in the room at all times.

3.1.3 Testing session

For the most part, the testing session equaled the baseline screening of the main study as described in section 2.3. It started with a measurement of the sEBR using EMG (ca. 15 min), followed by a computerized testing of WMC and discounting behavior (ca. 100 min), and ended with several

questionnaires (ca. 30 min). Note that the discounting tasks and questionnaires were conducted for another project and were not further analyzed in this study. The computerized testing in both sessions comprised eight tasks in random order. These included the six tasks already described for the baseline screening of the main study (see 2.3.2), plus two additional WMC tasks (Symmetry Span Task and Block Span Task), which are described in the following. For each task, the sequence of items to be recalled was different in the second (retest) session than in the first session.

The Symmetry Span Task was adopted from Foster et al. (2015) and belongs to the class of complex span tasks for measuring WMC (see Kane et al., 2004; Redick et al., 2012; Unsworth et al., 2009). The general procedure of this task was very similar to the Rotation Span Task and Operation Span Task described in section 2.3.2, except for the type of stimuli involved. In this task, subjects were required to memorize a sequence of square positions (memory component) while being distracted by a symmetry task (distraction component). The memory component of the task is described first. In each trial, a series of two to five small red squares was presented on screen, which were randomly positioned within a large 4x4 grid. Each red square appeared on screen for 650 ms, followed by a blank screen for 250 ms. Thereafter, subjects had to recall the positions of all red squares of the preceding series in the same order as they were presented. For recall, the 4x4 grid was displayed on screen and subjects were asked to click on the positions in which the red squares appeared in the preceding series in the correct order. Participants had no response time limit for recall. At the end of each trial, feedback about the number of correctly recalled red squares was presented on screen. In addition to this memory component, each trial also contained a distractor component. Before the presentation of each red square within a trial, subjects had to solve an item of a symmetry task. Each trial therefore contained as many symmetry task items as red squares (i.e. two to five), both being presented intermixed within one trial. In each item of the distractor task, a black-and-white picture appeared on screen, which was either left-to-right symmetric or not. Subjects were required to indicate as quickly as possible if the picture was left-to-right symmetric or not. About half of all pictures in the task were left-to-right symmetric. The response time limit for each subject was determined from their mean reaction time plus 2.5 standard deviations in the preceding practice block of the symmetry task (see below). If participants exceeded this limit, the picture disappeared and the item was counted as an error. After each item of the symmetry task, a blank screen appeared for 200 ms before the next red square was shown. At the end of each trial, subjects received feedback about their performance in the symmetry task (percent accuracy), averaged over all preceding trials. Participants were instructed to always maintain accuracy levels in the symmetry task above 85 %. In total, the Symmetry Span Task contained 12 trials, including three trials of each set size (two to five) in random order. Thus, 42 red squares and 42 pictures were presented in total. At the end of the task, two different memory scores were calculated: the partial score and the absolute score. The partial score equals the number of correctly recalled red squares out of all 42 presented red squares. The absolute score only includes the number of correctly recalled red squares within sets in which all red squares were recalled correctly.

In preparation for the task, participants completed three practice blocks. In the first practice block, only the memory component of the task was trained for four trials (two trials of set size two and two trials of set size three). In the second practice block, only the distractor component of the task was trained for 15 trials. After this block, the mean reaction time plus 2.5 standard deviations was calculated for each subject to determine the response time limit for this subject in the distractor component of the final task (see above). The third practice block trained the final task with both its memory and distractor component and contained three trials of set size two.

The Block Span Task was adapted from the Corsi block-tapping test (Corsi, 1972) and contained a forward and a backward version. In the original task, subjects are presented with a board with nine blocks in different spatial locations. The experimenter taps sequentially on some of the blocks and subjects need to recall the locations of the tapped blocks in the correct order (forward version) or in the reverse order (backward version). Note that the original task by Corsi (1972) only included the forward version, whereas the backward version was introduced later (see Berch, Krikorian, & Huha, 1998; Kessels, van den Berg, Ruis, & Brands, 2008). For the baseline screening, the task was adapted for PC in order to test multiple participants simultaneously. Here, the blocks were displayed on the screen in white color and were sequentially highlighted in blue color instead of being tapped by an experimenter. For recall, subjects clicked sequentially on the blocks rather than tapping them on the board. Except for these changes, the computerized version followed exactly the same procedure as the Corsi block-tapping test (described by Kessels et al., 2000, 2008). In the forward version, set sizes (i.e. sequence lengths) ranged from two to nine blocks and in the backward version from two to eight blocks. Both versions included two trials per set size, whereby set sizes were presented in ascending order. Each trial started with the presentation of a block sequence with a speed of one block per second. Thereafter, the German word “Erinnern” (recall) appeared on screen to prompt subjects to click on the blocks they recalled in the correct or reverse order, depending on the task version. There was no time limit for recall and subjects received no feedback about their memory performance. Participants needed to recall at least one trial per set size correctly to proceed to the next larger set size. The task terminated if both trials of the same set size were not recalled correctly. Thus, the number of trials depended on the subject’s performance with a maximum of 16 trials for the forward version and 14 trials for the backward version. At the end of the task, two different types of scores were calculated for each version: the span score and the total score. The span score is the size of the longest sequence recalled correctly in the normal order (span score forward) or in the reverse order (span score backward). The total score equals the total number of correct trials, calculated separately for the forward version (total score forward) and for the backward version (total score backward). The maximum achievable scores were nine for the span score forward, eight for the span score backward, 16 for the total score forward, and 14 for the total score backward. In preparation for the task, participants performed two practice trials (of set size two and three) for each version, in which they were given feedback about the correctness of their answer.

3.1.4 Data analysis

To quantify the one-week retest reliability of each instrument, the Pearson correlation coefficient was calculated between the task scores from both sessions over all subjects. Furthermore, the coefficient of variation (CV) was calculated for each measure on each session to compare the relative interindividual variability between the different instruments. The CV, also referred to as relative standard deviation, is defined as the standard deviation divided by the mean and is usually expressed in percent.

3.2 Study-specific results and conclusion

This pilot study examined the retest reliability and interindividual variability of different potential DA proxy measures, including the sEBR and different WMC task scores. Each variable was measured twice within the same sample (n=15) with a time interval of one week. The scores of both sessions are plotted against each other in Figure 6. For each measure, the retest reliability was quantified by the Pearson correlation coefficient and the relative interindividual variability by the CV, both shown in Table 2 (ordered by descending retest reliability). The Pearson correlation coefficient was significant for nearly all of the tested measures and ranged between .39 and .86. On the basis of these results, the five instruments with highest retest reliability were selected to be included in the baseline screening of the main study: the sEBR, the Rotation and Operation Span Task, the Listening Span Task, and the Digit Span Task. This selection also took into account that the total execution time of these tasks should not exceed 90 min to limit mental fatigue effects on cognitive performance (see Boksem, Meijman, & Lorist, 2005; Lorist, Boksem, & Ridderinkhof, 2005). Note that while the Digit Span Task was included with both its forward and backward version in the baseline screening of main study, only the latter may be considered a WMC measure due to its dual-task nature, whereas the forward version is rather a measure of simple short-term memory (see 1.2.5; e.g. Conway et al., 2005; Kail & Hall, 2001; Kane et al., 2004; Oberauer et al., 2000; Unsworth & Engle, 2007). Hence, only the Digit Span Backward was planned to be used as a DA proxy measure in the main study.

As some of the WMC tasks offered two alternative types of scoring, only the scoring method with the better retest reliability and CV was selected to be used in the main study (see Table 2). For the Digit Span Task, the total score was selected over the span score, since it showed a higher retest reliability and CV for both the forward and backward version. For the Rotation and Operation Span Task, the absolute score was selected over the partial score, since it showed considerably higher interindividual variability (CV) for both tasks and higher retest reliability for the Rotation Span Task, whereas retest reliabilities for the Operation Span Task were comparable between both scoring methods. Note, however, that these latter results diverge from findings of previous studies, which report higher retest reliabilities for partial over absolute scores in complex span tasks (see Redick et al., 2012).

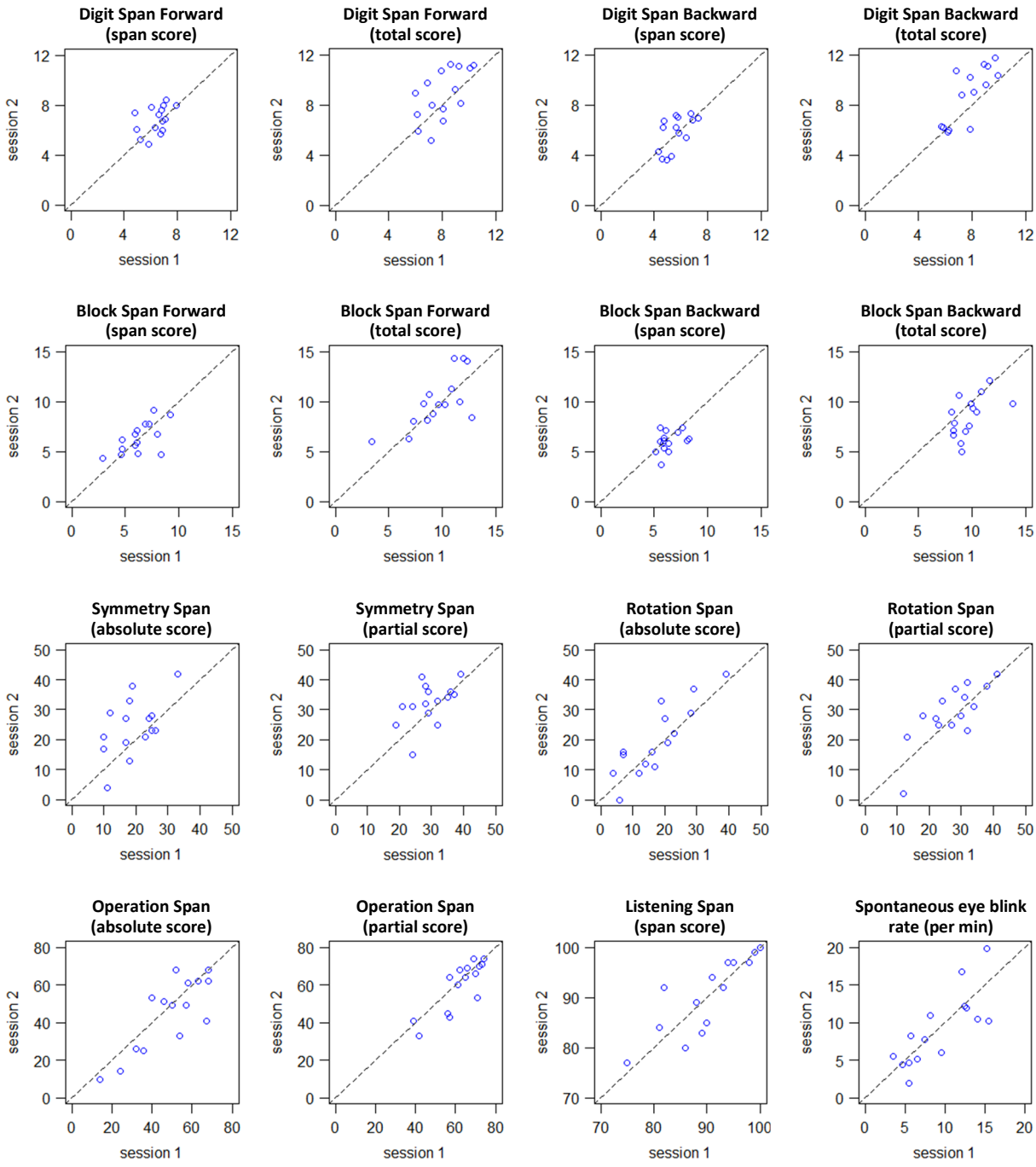


Figure 6. Scorings for all tested dopamine (DA) proxy measures. Each plot shows the scorings of both sessions plotted against each other, as well as a diagonal (dashed) line indicating equal values in both sessions.

Finally, it should be kept in mind that due to the small sample size, this pilot study only allowed for a rough estimation of the retest reliability and interindividual variability of the tested instruments. However, the retest reliabilities measured here are in line with the results of other studies with larger sample sizes, which also report values around .80 for the Digit Span Backward (Waters & Caplan, 2003), the Operation Span Task (Redick et al., 2012), and the sEBR (Barkley-Levenson & Galván, 2017; Dang et al., 2017; Kruis, Slagter, Bachhuber, Davidson, & Lutz, 2016; see also Jongkees & Colzato, 2016). Additionally, it should be considered that only some of these measures (e.g. the sEBR, Listening Span

Task, and Digit Span Task) have been used as DA proxies in previous studies (see 1.2.5). While it could be speculated that also the remaining WMC tasks tested in this pilot study should principally reflect central DA function, a direct link between these measures and central DA function has not yet been established by previous research. Note also that based on the results of the main study, a further discussion on the validity of these measures as proxies for DA function is provided in section 6.4.

In conclusion, five measures were selected to be used as DA proxies in the main study: the sEBR, the Rotation and Operation Span Task (absolute scores), the Listening Span Task (span score), and the Digit Span Backward (total score). All these measures showed adequate retest reliability (around .80; Carretero-Dios & Pérez, 2007; Nunnally & Bernstein, 1994) and mostly high interindividual variability.

Table 2. Retest reliabilities and coefficients of variation (CV) of all tested DA proxies.

Task (scoring method)	**Retest reliability**	**p value (Pearson)**	**CV (%)**	
			day1	**day2**
Rotation Span (absolute score) *	.86	< .001	56	59
Listening Span (span score) *	.84	< .001	8	8
Operation Span (partial score)	.83	< .001	17	22
Operation Span (absolute score) *	.81	< .001	34	43
Spontaneous eye blink rate (per min) *	.78	< .001	44	54
Digit Span Backward (total score) *	.78	< .001	19	26
Rotation Span (partial score)	.77	< .001	31	34
Block Span Forward (span score)	.74	.002	24	24
Digit Span Backward (span score)	.72	.004	15	23
Block Span Forward (total score)	.66	.007	27	26
Digit Span Forward (total score)	.64	.011	17	23
Block Span Backward (total score)	.58	.024	17	23
Symmetry Span (absolute score)	.54	.039	35	39
Digit Span Forward (span score)	.53	.041	14	16
Symmetry Span (partial score)	.51	.052	20	21
Block Span Backward (span score)	.39	.150	14	15

Note. * indicates measures that were selected as dopamine (DA) proxies for the main study.

4 Pilot study 2

The aim of this pilot study was to test the bandit task prior to its use in the main study and to compare different approaches for quantitative model comparison based on these data. Furthermore, some data of this pilot study were also included in the inverted-U analysis (see 2.9.2), which are hence reported and discussed later in the respective sections of the main study (see 5.5.1 and 6.4).

4.1 Study-specific methods

4.1.1 Participants

In total, 16 healthy male subjects participated in the study (aged 20 to 31 years, $M = 24.81$, $SD = 2.81$). Participants were recruited via an online job portal (www.stellenwerk.de) and were screened for the following inclusion criteria: male, age 18-35 years, right-handed, fluent German in speaking and writing, normal or corrected to normal vision, no hearing impairments, no major past or present psychological, neurological, or physical disorders, no regular consumption of prescription drugs. Before participating in the study, all subjects provided informed written consent, and study procedures were approved by the local ethics committee (Hamburg Medical Council). Participants were paid after the experiment with 10 € per hour plus variable monetary rewards from the bandit task (5-15 €), resulting in total payouts between 40 and 50 € per subject.

4.1.2 General procedure

Each subject was tested in one session, which started with the baseline screening as described for the main study (see 2.3), followed by the restless four-armed bandit task (see 2.4). The bandit task followed the same procedure as in the main study, except that it was performed outside the fMRI scanner in this pilot study. Sessions started between 9 am and 4 pm and lasted between 3.0 to 3.5 hours, separated by small breaks. Sessions were conducted in groups of two or four in the same experimental room, each subject sitting in front of a computer screen, separated by partition walls. An experimenter was present in the room at all times.

4.1.3 Cognitive modeling

For the most part, the cognitive modeling in this pilot study equaled the one in the main study (see 2.7.2). In short, choice behavior in the bandit task was modeled using six different cognitive models in a hierarchical Bayesian modeling approach. The six cognitive models were the same as in the main study, which resulted from the combination of two different learning rules (Delta rule, Bayesian learner) with three different choice rules (SM, SM+E, SM+EP). Bayesian modeling was performed with the software Stan (version 2.17.0; Stan Development Team, 2017b), operating from within the statistical package R (version 3.4.3; R Core Team, 2017) with the interface rstan (version 2.17.2; Stan Development Team, 2017a), using the same settings for sampling as in the main study (see 2.7.2).

Following parameter estimation, the cognitive models were compared in terms of their predictive accuracy using leave-one-out (LOO) cross-validation.

However, the cognitive modeling performed here differed from the one in the main study in two points. First, for the model comparison, exact LOO measures were calculated for each model in addition to the LOO estimates obtained by the R package loo (Vehtari et al., 2017). The general procedure of calculating exact LOO measures was already described in section 2.7.2. In short, for each subject in the data set (n=16), the model was fitted to a reduced data set excluding that subject (training set). Then, the predictive accuracy of the model was calculated for the left-out subject (testing set), using the log likelihood function evaluated at the sampled posterior parameter values for that subject, which were obtained from fitting the model to the testing set. To obtain subject-specific posteriors for the left-out subject in these model fits, these fits included one additional subject without any data points, for which posterior samples were drawn. Without any data points included for this additional subject and with uninformative priors, this subject's posterior samples were only informed by the posterior samples for the group-level parameters of the hierarchical model (M^x and Λ^x for any parameter x, see Figure 5), which are in turn informed by the data in the training set. Finally, the averaged log likelihoods (i.e. averaged across posterior samples) for all subjects were then summed to yield the overall LOO measure for each cognitive model (as described by Vehtari et al., 2017).

Second, in addition to the exact and estimated LOO measures, an alternative measure of predictive accuracy was used for the model comparison in this pilot study, which was calculated over left-out trials and not over left-out subjects. In the following, this measure is denoted as CV_{trials} (for cross-validation over trials). For the CV_{trials} measure, the choice data of each subject were divided into a training set, containing the first 240 trials of the bandit task, and a testing set, containing the last 60 trials. Each cognitive model was then fitted to the training set to obtain subject-specific parameter posteriors based only on the first 240 trials. Next, it was examined how well each of the fitted models predicted choices in the last 60 trials, using the log likelihood function evaluated at the sampled posterior parameter values obtained from the training set, summed over all 60 trials of the testing set. Finally, the averaged log likelihoods for all subjects were then summed to yield the overall CV_{trials} measure for each cognitive model (see Vehtari et al., 2017).

After model comparison, the winner model Bayes-SM+EP was fitted again to the data of this pilot study, but this time treating the random walk parameters $\hat{\sigma}_d^2$, $\hat{\lambda}$, $\hat{\vartheta}$, $\hat{\mu}_1^{pre}$, and $\hat{\sigma}_1^{pre}$ as free parameters in order to use their posterior estimates (medians) in the main study (see 2.7.2). Each of these parameters was estimated once over all 16 subjects, using uniform priors within the following limits: $[0, \infty]$ for $\hat{\sigma}_d^2$; $[0, 1]$ for $\hat{\lambda}$; $[0, 100]$ for $\hat{\vartheta}$; $[0, 100]$ for $\hat{\mu}_1^{pre}$; $[0, \infty]$ for $\hat{\sigma}_1^{pre}$.

4.1.4 Data analysis

Several data from this pilot study, including data from the baseline screening and the bandit task, were analyzed in combination with data from the main study to increase the sample size for these analyses. Procedures for these combined data analyses are described in the respective methods sections of the main study, i.e. section 2.7.2 for a correlation analysis between the model-based choice parameters and the percentage of exploratory trials, and section 2.9.2 for the inverted-U analysis.

4.2 Study-specific results and conclusion

In this pilot study, six cognitive models of learning and decision making were compared using different cross-validation approaches. First, all models were fitted to the data (n=16) in a hierarchical Bayesian modeling framework and then compared using three measures of predictive accuracy: an exact leave-one-out measure (LOO exact), an estimated leave-one-out measure using the R package loo (LOO estimate), and a cross-validation measure calculated over left-out trials instead of left-out subjects (CV_{trials}). The results of all predictive accuracy measures for all six cognitive models are shown in Table 3 and plotted in Figure 7. Note that Figure 7 shows predictive accuracies per subject and trial to allow for a better visual comparison of the three measures, which were compounded over different numbers of trials. Notably, all three approaches show the same winner model (Bayes-SM+EP), despite slight variations in the order of the remaining five models. Since LOO estimates yielded comparable results to the alternative measures, but are much faster to compute, they were subsequently selected as the measure for model comparison in the main study. Since the model comparison results of this pilot study largely equaled those of the main study, a more detailed discussion of these findings is provided in the respective section of the main study (see 6.5).

Table 3. Results of the cognitive model comparison in pilot study 2.

cognitive model	LOO estimate	order	LOO exact	order	CV_{trials}	order
Delta-SM	-3559.8	6	-3576.5	6	-778.2	5
Bayes-SM	-3435.9	5	-3453.7	4	-711.0	3
Delta-SM+E	-3431.8	4	-3468.9	5	-803.3	6
Bayes-SM+E	-3334.7	3	-3368.2	3	-707.3	2
Delta-SM+EP	-3235.2	2	-3350.6	2	-714.7	4
Bayes-SM+EP	-3193.8	1	-3243.0	1	-670.3	1

Note. Three different cross-validation measures were calculated: two leave-one-out (LOO) measures over subjects, either calculated exactly (LOO exact) or estimated with the R package loo (LOO estimate), and one cross-validation measure over trials (CV_{trials}). SM: softmax; SM+E: softmax with exploration bonus; SM+EP: softmax with exploration bonus and perseveration bonus.

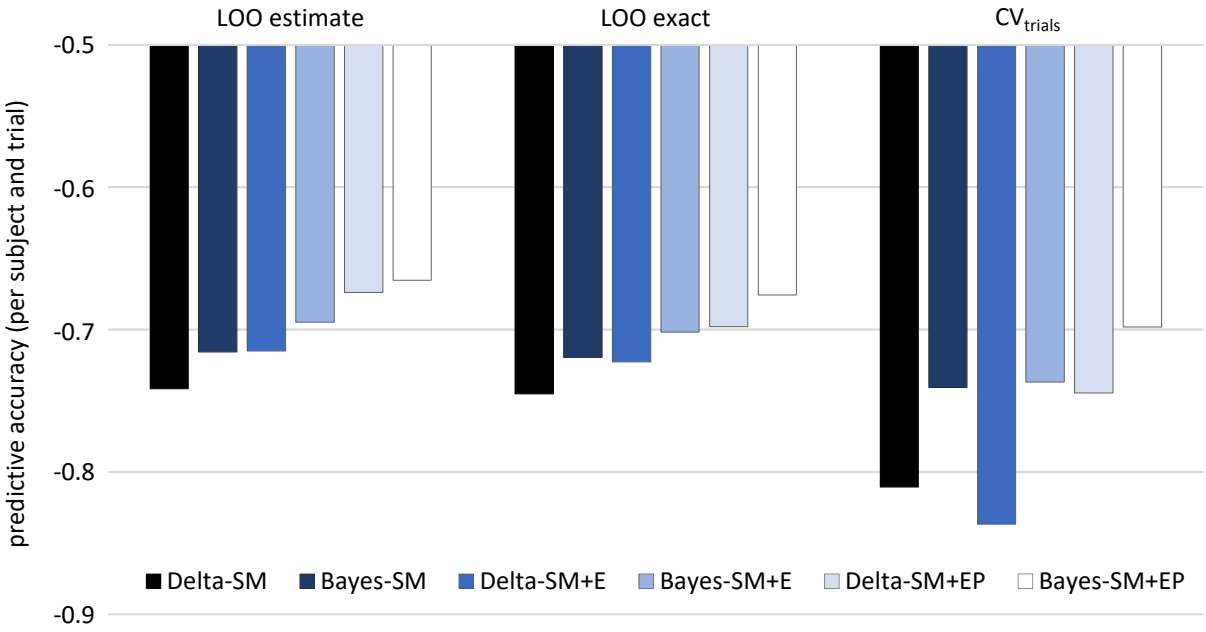


Figure 7. Results of the cognitive model comparison in pilot study 2. Three different cross-validation measures were calculated: two leave-one-out (LOO) measures over subjects, either calculated exactly (LOO exact) or estimated with the R package loo (LOO estimate), and one cross-validation measure over trials (CV_{trials}). All cross-validation measures were devided by the total number of data points in the sample (n*t with n=16 subjects, t=300 trials for LOO, and t=60 trials for CV_{trials}) for better comparability across the different approaches. Note that such linear transformations do not alter the relative order of the cross-validation measures for the six cognitive models, which is of main relevance for the model comparison. SM: softmax; SM+E: softmax with exploration bonus; SM+EP: softmax with exploration bonus and perseveration bonus.

5 Main results

5.1 Cognitive model comparison

After the initial model comparison in pilot study 2 (see 4.2), the six cognitive models were compared again based on the larger sample of the main study. First, all cognitive models were fitted separately to each drug condition using hierarchical Bayesian modeling. Then, LOO estimates were calculated as a measure of predictive accuracy, once separately for each drug condition and once compounded over all three drug conditions. LOO estimates for all six cognitive models are shown in Table 4 and plotted in Figure 8. Note that Figure 8 shows predictive accuracies per subject and trial to allow for a better visual comparison of all measures (and to the results of pilot study 2), which were compounded over different numbers of trials. Altogether, the model comparison over drugs and the three drug-specific model comparisons consistently showed the same winner model, Bayes-SM+EP (as in pilot study 2; see 4.2), as well as the same overall order of all six cognitive models.

Of particular interest is the finding that the novel Bayes-SM+EP model with an extra perseveration bonus parameter outperformed the Bayes-SM+E model, which does not account for perseveration. It has been argued in the literature that perseveration, if not explicitly accounted for in the cognitive model, might be captured by the exploration bonus parameter and interpreted as an uncertainty-avoiding choice bias (as discussed in section 6.5.2; see Badre et al., 2012; Payzan-LeNestour & Bossaerts, 2012). Hence, one question of interest was how estimates for the exploration bonus parameter (φ) were affected by introducing the extra perseveration bonus parameter (ρ) into the model. Therefore, φ parameter estimates for the placebo condition were compared between the winner model Bayes-SM+EP (with perseveration bonus) and the Bayes-SM+E model (without perseveration bonus). The comparison showed that subject-level φ medians correlated highly between both models ($r_{29} = .90$, $p < .001$), but were significantly higher for the Bayes-SM+EP than for the Bayes-SM+E model (mean difference = 0.79, paired t-test: $t_{30} = 7.97$, $p < .001$). The number of subjects who showed a negative φ median, reflecting a discouragement rather than an encouragement of uncertainty-driven exploration, was 13 of 31 for the Bayes-SM+E model, but only 6 of 31 for the Bayes-SM+EP model. Also, the corresponding group-level mean parameter M^{φ} was considerably higher for the Bayes-SM+EP model (M^{φ} median = 0.95) than for the Bayes-SM+E model (M^{φ} median = 0.16).

After model selection, the winner model Bayes-SM+EP was fitted again to the data, but this time the five random walk parameters of the Bayesian learner rule were fixed to their posterior medians estimated over all subjects of pilot study 2 (see 2.7.2 and Table A1 in the appendix) instead of fixing them arbitrarily to the true values of these parameters. Note that fixing the random walk parameters to their true values was done only initially in order to include all six models in the model comparison, since some of these models did not converge with free random walk parameters. Predictive accuracy of this new model fit proved to be even better (smaller absolute LOO values) than for the previous

model fit, as shown in Table 4. Thus, this improved Bayes-SM+EP model fit, in which the random walk parameters were fixed to their posterior medians from pilot study 2, was selected to be used for all further model-based analyses.

Table 4. Results of the cognitive model comparison in the main study.

cognitive model	LOO over drugs	LOO placebo	LOO L-dopa	LOO haloperidol	order (all)
Delta-SM	-17698.7	-6088.1	-5724.7	-5897.3	6
Bayes-SM	-17428.5	-6037.7	-5590.4	-5807.4	5
Delta-SM+E	-17220.2	-5932.7	-5513.8	-5798.0	4
Bayes-SM+E	-16942.9	-5891.2	-5386.5	-5686.5	3
Delta-SM+EP	-16724.4	-5782.5	-5349.8	-5610.9	2
Bayes-SM+EP	-16269.9	-5652.2	-5182.7	-5465.4	1
Bayes-SM+EP[a]	-15546.9	-5356.9	-4987.7	-5232.7	-

Note. LOO estimates were calculated once over all drug conditions (column: over drugs) and once separately for each drug condition (columns: placebo, L-dopa, haloperidol). SM: softmax; SM+E: softmax with exploration bonus; SM+EP: softmax with exploration bonus and perseveration bonus.

[a] Random walk parameters for this model fit were fixed to their posterior medians from pilot study 2, not to their true values.

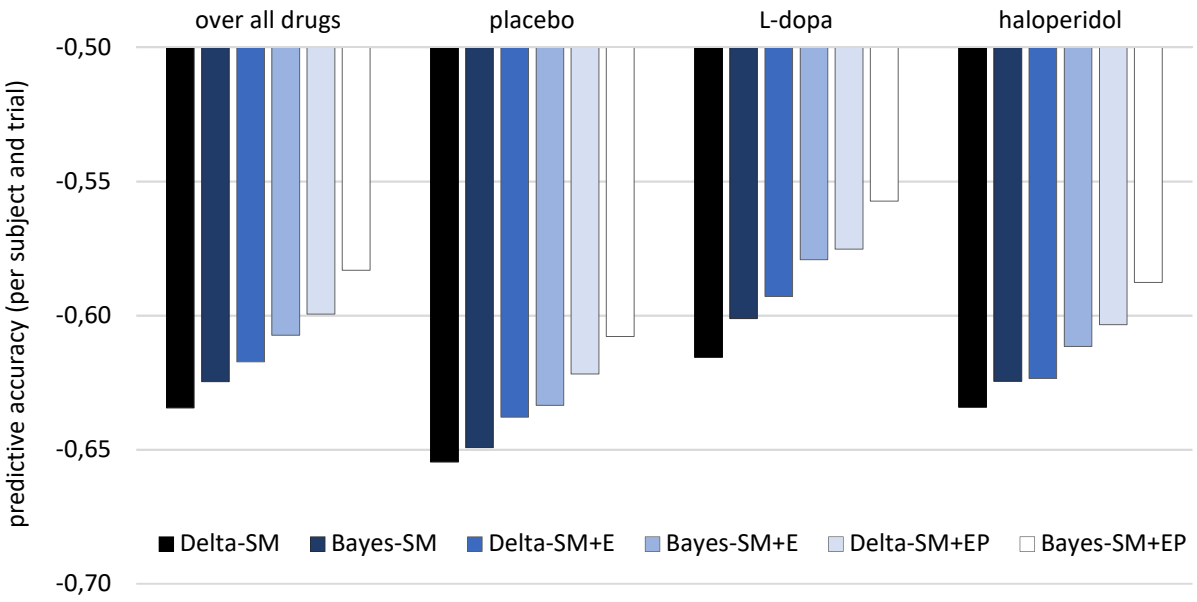


Figure 8. Results of the cognitive model comparison in the main study. Leave-one-out (LOO) estimates were calculated once over all drug conditions (n=31 with t=3*300) and once separately for each drug condition (n=31 with t=300). All LOO estimates were devided by the total number of data points in the sample (n*t) for better comparability across the different approaches. Note that such linear transformations do not alter the relative order of LOO estimates for the six cognitive models, which is of main relevance for the model comparison. SM: softmax; SM+E: softmax with exploration bonus; SM+EP: softmax with exploration bonus and perseveration bonus.

5.2 Model-based behavioral results

5.2.1 Bayes-SM+EP model characterization

Since all model comparison approaches of the main study and pilot study 2 consistently resulted in the same winner model, Bayes-SM+EP, this model was selected for further analyses of the behavioral and fMRI data. To facilitate the understanding and interpretation of these model-based analyses and their results, some characteristics of the Bayes-SM+EP model should be considered first. Therefore, the most relevant trial-by-trial quantities of this model are plotted in Figure 9, derived from the placebo data of one representative subject (with the following posterior medians: $\beta = 0.29$, $\varphi = 1.34$, and $\rho = 4.11$). According to this model, subjects' choices are stochastically dependent on three factors: the expected reward value ($\hat{\mu}^{pre}$; Figure 9a), the exploration bonus ($\varphi\hat{\sigma}^{pre}$; Figure 9b), and the perseveration bonus ($I\rho$; Figure 9c) of each bandit. Based on these quantities, the model predicts the choice probability for each of the four bandits on each trial (P; Figure 9d). Crucially, the size of the φ and ρ parameter determine how these three factors are weighted for the choice prediction, i.e. how strongly choices are driven by uncertainty and perseveration relative to the bandits' expected reward values. Another important quantity of this model is the reward prediction error (δ; Figure 9e), which is the difference between the received and expected reward on a given trial ($r - \hat{\mu}^{pre}$; both shown in Figure 9a) and which is used to adjust value predictions for the selected bandit on the following trial. Note also how the expected values of all unchosen bandits drift from trial to trial towards a certain value (here: 45.99), a behavior that is determined by the random walk parameters $\hat{\lambda}$ (decay) and $\hat{\vartheta}$ (decay center) of the Bayes-SM+EP model. This decay may capture a subject's belief that extreme reward values drift back towards a central (or the initial) value, but also a gradual forgetting of expected reward values over time. Finally, another quantity, which was specifically of interest for the model-based fMRI analysis, is the summed uncertainty of all four bandits ($\Sigma\hat{\sigma}^{pre}$; Figure 9f), which was employed as a measure for the subject's overall uncertainty in a given trial.

Furthermore, it is important to note that these trial-by-trial quantities can be used to classify subjects' choices into different choice types. According to Daw et al. (2006), choices can be classified in a binary fashion as either exploitations, i.e. choosing the bandit with the highest expected value, or explorations, i.e. choosing one of the other bandits. For demonstration, all exploration trials (according to this binary classification) are marked by a black line in the lower half of Figure 9a. However, this classification defines exploration rather broadly as all choices that are not value-driven, but does not further distinguish between different types of exploration (i.e. directed vs. random exploration). Therefore, a second classification scheme was used here, which divides choices into exploitations, directed explorations, and random explorations. According to this trinary classification, exploitations are all trials in which the bandit with the highest expected value is chosen, or the bandit with the highest sum of expected value plus perseveration bonus. All the remaining (i.e. non-exploitative) trials are classified as either directed explorations, i.e. trials in which the bandit with the highest exploration bonus was chosen, or random explorations, i.e. trials in which not the bandit with the highest

exploration bonus was chosen. Note, however, that according to this trinary classification, "random explorations" can also be driven partly (but not solely) by a bandit's expected value or exploration bonus.

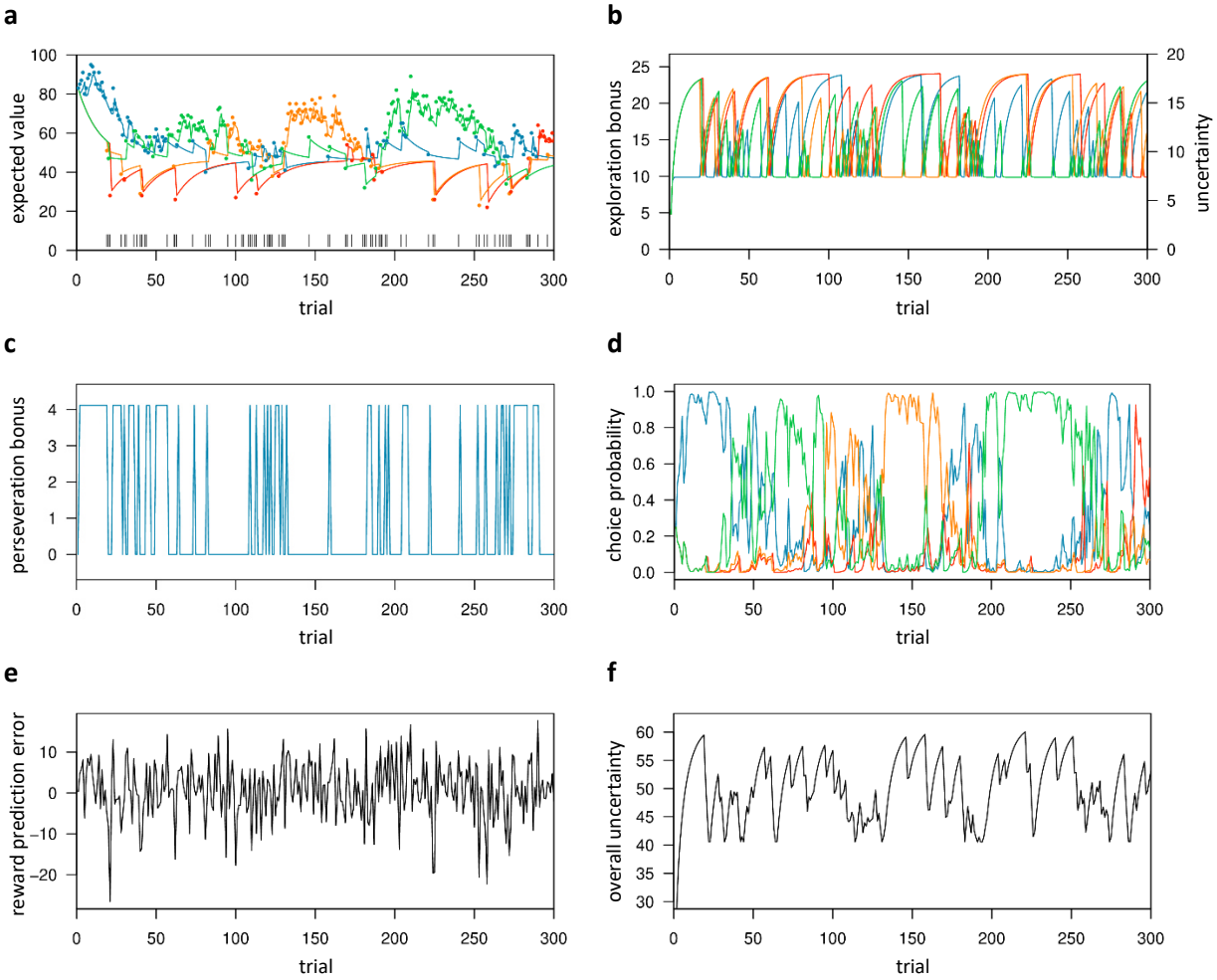


Figure 9. Trial-by-trial variables of the Bayes-SM+EP model. Trial-by-trial estimates were derived from the placebo data of one representative subject with the following posterior medians: $\beta = 0.29$, $\varphi = 1.34$, and $\rho = 4.11$. **(a)** Expected value ($\hat{\mu}^{pre}$) and actual payoff (r). Each colored line shows the expected value of one bandit for all trials, whereas colored dots mark the size of the actual payoffs on all trials. Black lines in the lower part of the plot mark exploratory trials according to the binary classification by Daw et al. (2006). **(b)** Exploration bonus ($\varphi\hat{\sigma}^{pre}$) and uncertainty ($\hat{\sigma}^{pre}$). Each colored line represents one bandit. Note that the exploration bonus is merely the uncertainty scaled by the φ parameter in order to express the bonus in value units. **(c)** Perseveration bonus ($I\rho$). This bonus is a fixed value added only to the bandit chosen in the previous trial, shown here for one bandit. **(d)** Choice probability (P). Each colored line represents one bandit. **(e)** Reward prediction error *(δ).* **(f)** Overall uncertainty ($\Sigma\hat{\sigma}^{pre}$). This plot shows the summed uncertainty ($\hat{\sigma}^{pre}$) over all four bandits, which was calculated as a measure of the subject's overall uncertainty in a given trial.

Based on these classifications, it was evaluated how these different choice types related to the three choice parameters (β, φ, ρ) of the Bayes-SM+EP model. Therefore, the percentage of overall explorations (according to the binary classification) as well as the percentage of random and directed explorations (according to the trinary classification) were calculated for each subject over all trials per session, and Pearson correlations between these percentages and the subject-specific posterior

medians of each choice parameter were computed (see Table 5). Note that data from the placebo condition (n=31) and pilot study 2 (n=16) were combined to increase the sample size for this correlation analysis (n=47). When using the binary classification, the percentage of explorations per subject correlated negatively with the β parameter, but positively with the φ parameter. However, when further subdividing exploratory choices according to the trinary classification, the percentage of directed explorations showed a significant positive correlation only with the φ parameter, while the percentage of random explorations showed a significant negative correlation only with the β parameter, consistent with the idea that both parameters should reflect different types of exploration (see 1.1.3 and 2.7.2).

Table 5. Correlations between different exploration types and model-based choice parameters.

% explorations	β	φ	ρ
overall	-.65***	.30*	-.18
directed	-.28	.64***	-.09
random	-.68***	-.09	-.22

Note. Reported are Pearson correlation coefficients between the percentage of exploratory trials per subject and the subject-level parameter medians of the Bayes-SM+EP model. Note that overall explorations were defined according to the binary choice classification, while directed and random explorations were defined according to the trinary choice classification. β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter.

*$p < .05$. ***$p < .001$.

5.2.2 Drug effect analysis on group-level parameters

Dopaminergic drug effects were first analyzed for the group-level parameters of the Bayes-SM+EP model, which are the mean (M) and standard deviation (Λ) for each of the three choice parameters (β, φ, ρ). Posterior densities for these six group-level parameters were estimated separately for each drug condition and are plotted in Figure 10. Each plot shows the posterior median (vertical black line) as well as the 80% central interval (i.e. from the 10th to 90th percentile; blue area) and 95% central interval (i.e. from the 2.5th to 97.5th percentile; black contours) of the posterior distribution. These plots show a clear reduction of the group-level mean parameter for φ (M^{φ}) under L-dopa compared to placebo and haloperidol, as well as a reduction of the group-level standard deviation for φ (Λ^{φ}) under haloperidol compared to placebo and L-dopa. For better visualization of these drug effects, the pairwise drug differences of the M^{φ} and Λ^{φ} posterior densities are plotted in Figure 11. Additionally, Table 6 shows for each of these posterior differences the percentage of samples with values above zero and the 90% HDI. Notably, for the M^{φ} parameter, 97.4% of the posterior difference for the comparison placebo minus L-dopa lay above zero, while 98.3% of the posterior difference for the comparison L-dopa minus haloperidol lay below zero. For both of these M^{φ} posterior differences, zero was not included in the 90% HDI.

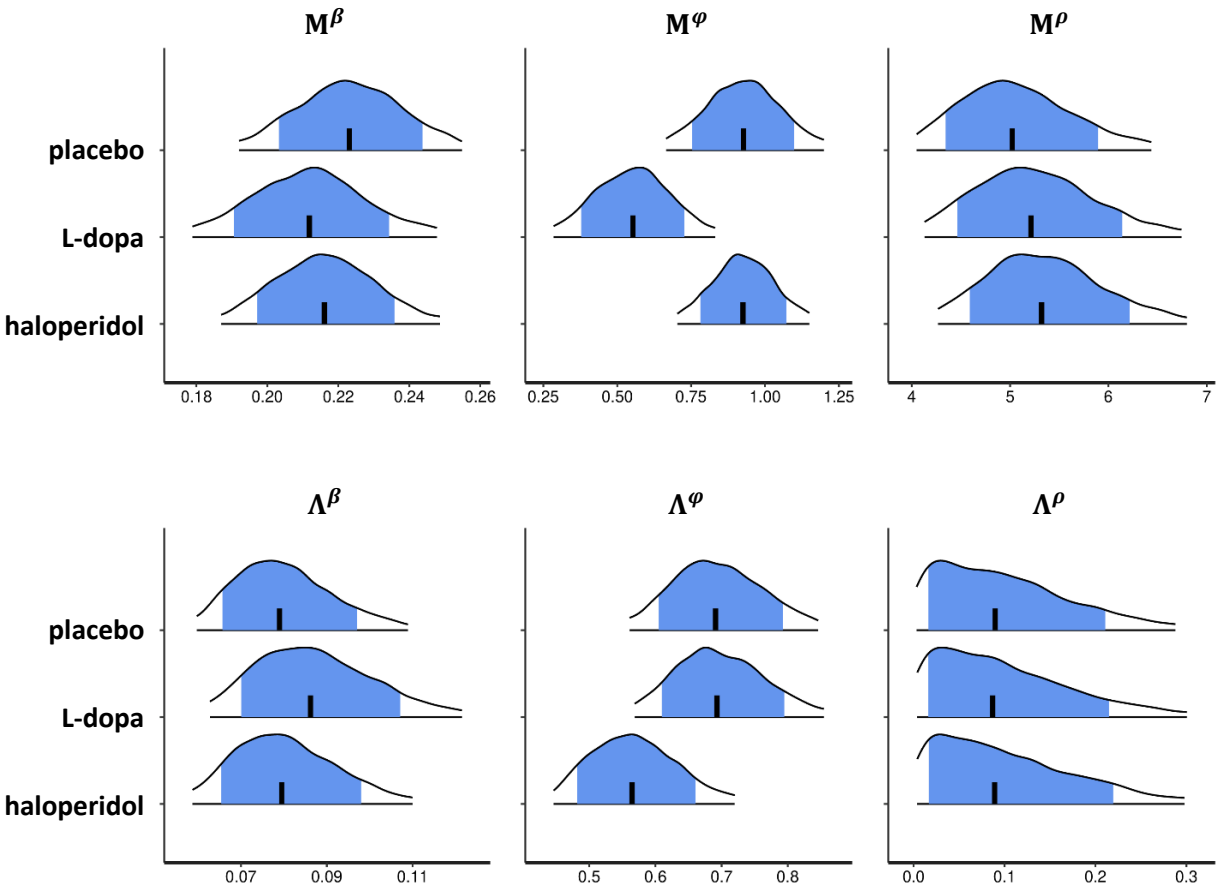


Figure 10. Group-level parameter estimates of the Bayes-SM+EP model. Shown are posterior distributions of the group-level mean (M) and standard deviation (Λ) for all choice parameters (β, φ, ρ) of the Bayes-SM+EP model, separately for each drug condition. For each posterior distribution, the plot shows the median (vertical black line), the 80 % central interval (blue area), and the 95 % central interval (black contours). β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter.

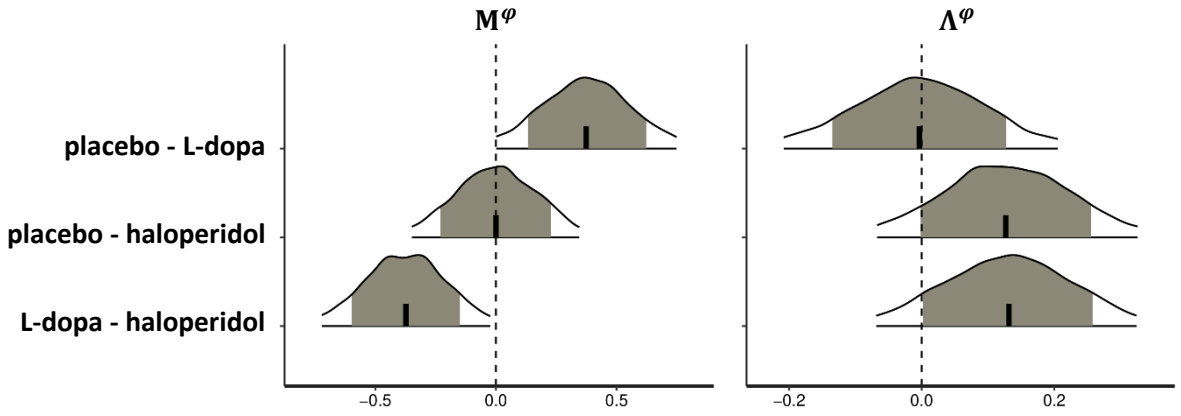


Figure 11. Drug effects on the exploration bonus parameter (φ) on the group level. Shown are posterior drug differences of the group-level mean (M) and standard deviation (Λ) for the φ parameter of the Bayes-SM+EP model. For each posterior distribution, the plot shows the median (vertical black line), the 80 % central interval (grey area), and the 95 % central interval (black contours).

Table 6. Drug effects on the exploration bonus parameter (φ) on the group level.

	M^{φ}		Λ^{φ}	
	% above 0	90 % HDI	% above 0	90 % HDI
placebo - L-dopa	97.5	[0.05, 0.69]	47.5	[-0.18, 0.16]
placebo - haloperidol	49.3	[-0.30, 0.27]	90.0	[-0.04, 0.29]
L-dopa - haloperidol	1.7	[-0.70, -0.10]	90.8	[-0.02, 0.31]

Note. Results refer to the posterior drug differences of the group-level mean (M^{φ}) and standard deviation (Λ^{φ}) for the φ parameter of the Bayes-SM+EP model. For each posterior difference, the table shows the percentage of samples with values above zero (column: % above 0) and the 90 % highest density interval (column: 90 % HDI).

5.2.3 Drug effect analysis on subject-level parameters

Next, dopaminergic drug effects were analyzed for the subject-level choice parameters (β, φ, ρ) of the Bayes-SM+EP model. Posterior densities for these parameters were estimated separately for each drug condition in a hierarchical design (see below for results on non-hierarchical designs) and are plotted in Figure 12 for the φ parameter and in Figure A1 and A3 of the appendix for the β and ρ parameter, respectively. Note that for each parameter, posterior densities are ordered ascendingly by their median in the placebo condition to better visualize their relative changes in the L-dopa and haloperidol condition. In addition, pairwise drug differences of these posterior densities are plotted in Figure 13 for the φ parameter and in Figure A2 and A4 of the appendix for the β and ρ parameter, respectively. Results for the φ parameter are reported first. From visual inspection of Figure 12 and 13, an overall trend towards a reduction of φ under L-dopa compared to placebo and haloperidol was observed, although some of the subjects also showed the opposite or no clear drug effect under L-dopa. Such an overall trend could not be observed under haloperidol, although visual inspection pointed towards the pattern that haloperidol increased φ for subjects with a relatively low φ value under placebo and decreased it for subjects with a relatively high φ value under placebo. In accordance with this, it was found that haloperidol reduced the variability of the subject-level φ medians (range = -0.43 to 2.33; SD = 0.64) compared to placebo (range = -0.95 to 2.48; SD = 0.85) and L-dopa (range = -1.77 to 2.00; SD = 0.85). For the β and ρ parameter, visual inspection of the respective plots (Figure A1 to A4 in the appendix) showed no systematic trend or pattern between drug conditions over all subjects, although drug effects of different magnitude and direction were present on the individual level.

To test for drug effects on the subject-level parameters, a repeated measures ANOVA with the factor drug and with posterior medians as the dependent variable was performed for each of the three choice parameters, followed by paired t-tests. Results of the ANOVA and t-tests are summarized in Table 7. The ANOVA showed a significant drug effect only for the φ parameter ($F_{2,60}$ = 4.54, p = .015), but not for the β and ρ parameter. The paired t-tests revealed a significant difference of the φ parameter for the comparisons placebo vs. L-dopa (t_{30} = 2.81, p = .009) and L-dopa vs. haloperidol (t_{30} = -2.34, p = .026), but not for the comparison placebo vs. haloperidol (t_{30} = -0.01, p = .991).

Next, the robustness of this L-dopa effect against different modeling approaches was tested. Therefore, the paired t-test comparison of the subject-level φ medians between the placebo and L-dopa condition was repeated based on the parameter estimates from the following alternative model fits:

(1) the Bayes-SM+E model without the perseveration bonus parameter ($t_{30} = 3.15$, $p = .004$);
(2) the Delta-SM+EP model with the Delta learner instead of the Bayesian learner ($t_{30} = 2.97$, $p = .006$);
(3) the Bayes-SM+EP model fitted in a non-hierarchical design, i.e. without the group-level parameters M and Λ ($t_{30} = 2.11$, $p = .043$);
(4) the Bayes-SM+EP model, in which all random walk parameters were fixed to their true values (see 2.4 and 2.7.2) instead of their posterior medians derived from pilot study 2 ($t_{30} = 3.44$, $p = .002$);
(5) the Bayes-SM+EP model with free random walk parameters, i.e. each random walk parameter was estimated once over all subjects (as described in 4.1.3) for each drug condition ($t_{30} = 4.20$, $p < .001$).

Note that with the last modeling approach, a direct comparison of φ estimates between drugs conditions is more complex in its interpretation, since each drug condition exhibits different random walk parameters (see Table A1 in the appendix). Taken together, all of these modeling approaches showed a significant ($p < .05$) reduction of the φ parameter under L-dopa compared to placebo.

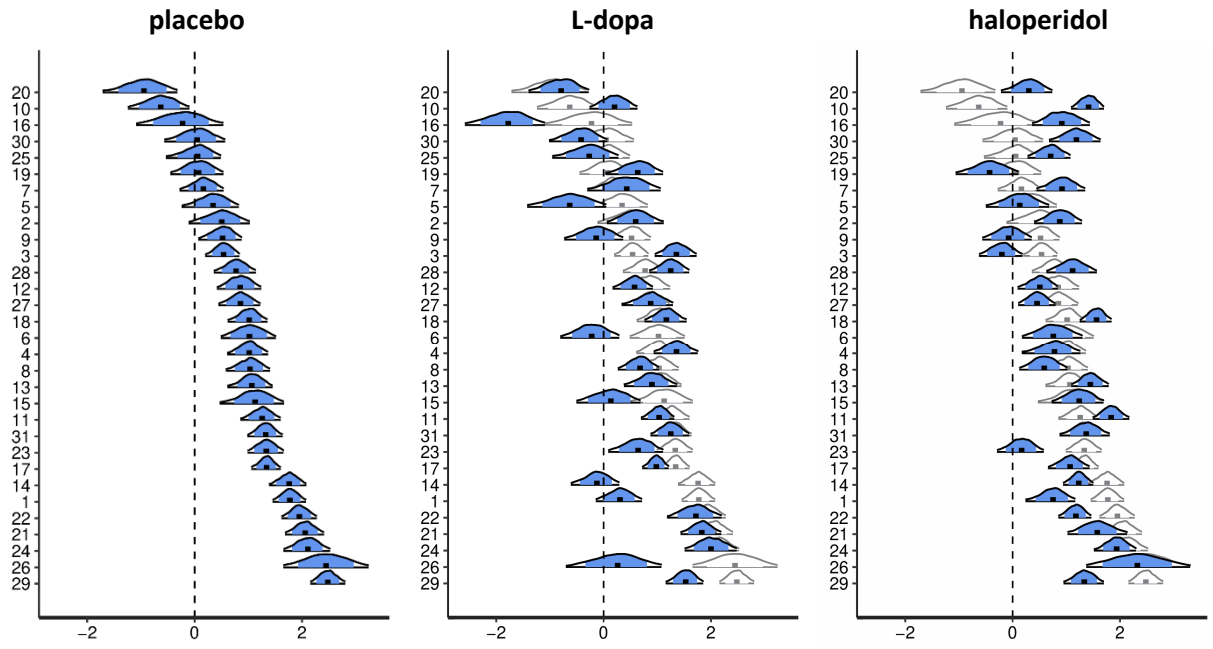


Figure 12. Subject-level parameter estimates for the exploration bonus parameter (φ). Shown are posterior distributions of the subject-level φ parameter of the Bayes-SM+EP model, separately for each drug condition. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (blue area), and the 95 % central interval (black contours). For the L-dopa and haloperidol condition, posterior distributions (in blue) are overlaid on the posterior distributions of the placebo condition (in white) for better comparison.

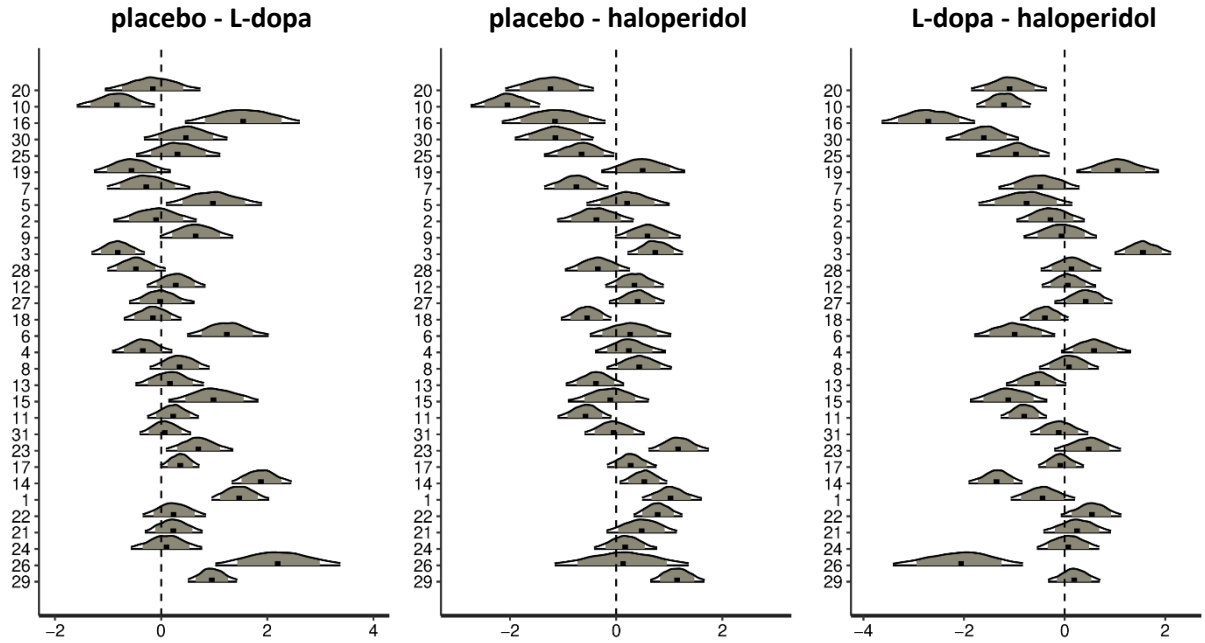


Figure 13. Drug effects on the exploration bonus parameter (φ) on the subject level. Shown are posterior drug differences of the subject-level φ parameter of the Bayes-SM+EP model. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (grey area), and the 95 % central interval (black contours).

Table 7. Testing for drug effects on the subject-level choice parameters.

	ANOVA		t-test: P - D			t-test: P - H			t-test: D - H		
	$F_{2,60}$	p	diff	t_{30}	p	diff	t_{30}	p	diff	t_{30}	p
β	0.34	.714	0.011	0.83	.414	0.007	0.52	.606	-0.004	-0.29	.776
φ	4.54	.015	0.377	2.81	.009	-0.002	-0.01	.991	-0.378	-2.34	.026
ρ	1.81	.172	0.613	0.73	.469	-0.818	-1.48	.149	-1.431	-1.71	.098

Note. The subject-level posterior medians of the three choice parameters of the Bayes-SM+EP model were used as dependent variables for the repeated measures ANOVA and paired t-tests. β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter; P: placebo; D: L-dopa; H: haloperidol; diff: mean difference.

5.2.4 Drug effect analysis on the percentage of exploitations and explorations

In addition to the three choice parameters, drug effects were also analyzed for the percentage of exploitative and exploratory trials per subject, which were determined according to different classification schemes (see 2.7.2). For this, a repeated measures ANOVA with the factor drug was performed for each of the following four dependent variables: (a) the percentage of overall explorations, (b) the percentage of directed explorations, (c) the percentage of random explorations, and (d) the percentage of exploitations, with (a) following the binary classification scheme by Daw et al. (2006), and (b) to (d) following the trinary classification scheme. Note that results for the percentage of exploitations according to the binary classification scheme are not explicitly reported, since these can be derived from the results for (a) by sign reversal. The corresponding data are plotted in Figure 14. The ANOVA showed a significant drug effect only for the percentage of directed explorations

($F_{2,60} = 7.15$, $p = .002$), but not for the percentage of random explorations ($F_{2,60} = 0.53$, $p = .592$), overall explorations ($F_{2,60} = 0.97$, $p = .386$), or exploitations ($F_{2,60} = 1.62$, $p = .207$). Next, paired t-tests were performed to compare the percentage of directed explorations pairwise between the three drug conditions. The t-tests showed a significant reduction in the percentage of directed explorations under L-dopa compared to placebo (mean difference P-D = 2.82, $t_{30} = 4.69$, $p < .001$) and haloperidol (mean difference H-D = 2.42, $t_{30} = 2.76$, $p = .010$), but not between placebo and haloperidol (mean difference P-H = 0.39, $t_{30} = 0.43$, $p = .667$). Note also that a further exploratory t-test revealed that the percentage of exploitations was marginally increased under L-dopa compared to placebo (mean difference P-D = -2.61, $t_{30} = -1.92$, $p = .065$).

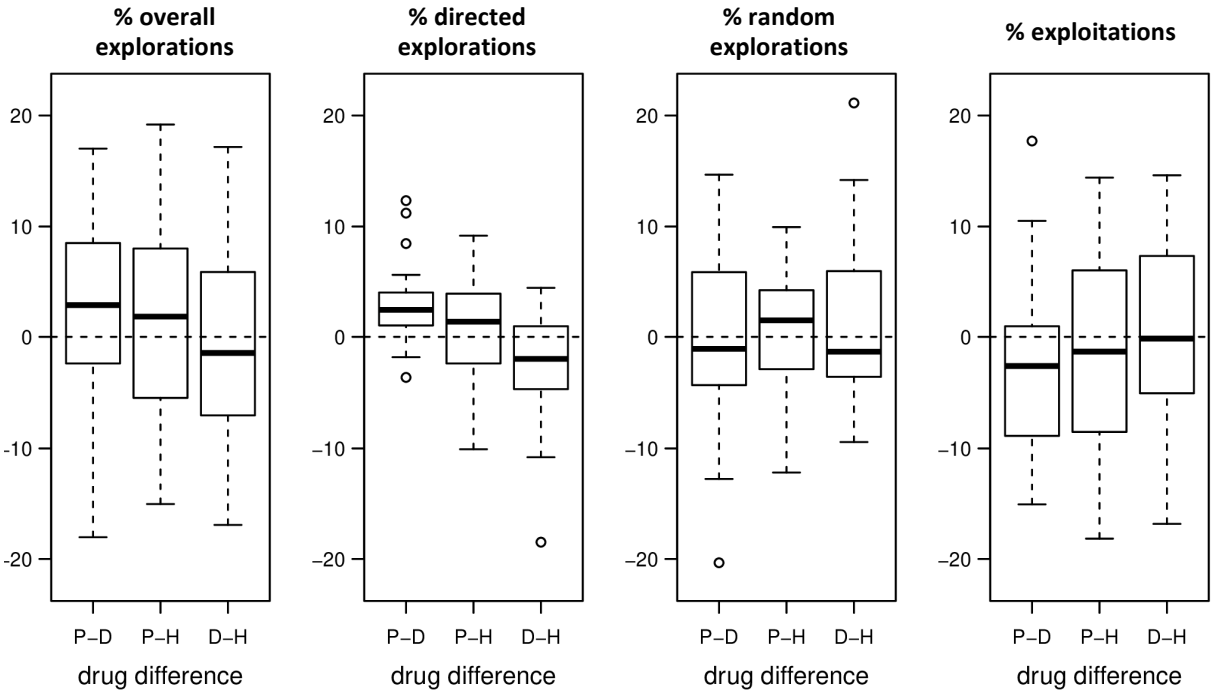


Figure 14. Drug effects on the percentage of explorations and exploitations. Shown are the pairwise drug differences for the percentage of overall explorations, directed explorations, random explorations, and exploitations. Note that overall explorations were defined according to the binary choice classification by Daw et al. (2006), whereas the other three choice types were defined according to the trinary choice classification. P: placebo; D: L-dopa; H: haloperidol.

5.3 Model-free behavioral results

In addition to the model-based choice parameters, also some model-free measures of choice behavior were tested for DA drug effects. These variables were the overall monetary payout (*payout*), the percentage of choices in which the bandit with the highest actual payoff was selected (*% best bandit*), the mean rank of all chosen bandits when bandits are ranked by their actual payoff (*mean rank*), and the percentage of switches (*% switches*). Yet, the repeated measures ANOVA yielded no significant drug effect on any of these four model-free choice variables (*payout*: $F_{2,60} = 0.06$, $p = .943$; *% best bandit*: $F_{2,60} = 0.34$, $p = .711$; *mean rank*: $F_{2,60} = 0.37$, $p = .690$; *% switches*: $F_{2,60} = 1.02$, $p = .366$).

5.4 Control variables

Several behavioral control variables were tested for DA drug effects. The first set of control variables was measured during the post-fMRI testing and included the spontaneous eye blink rate (sEBR), the total scores of the Digit Span Task (forward and backward), and 15 attentional performance measures from the Tests of Attentional Performance (TAP). To test for DA drug effects, a repeated measures ANOVA with the factor drug was performed on each of these 18 control variables. None of these 18 variables showed a significant drug effect in the ANOVA (all $p > .05$). The complete ANOVA results for the first set of control variables can be found in Table A2 of the appendix.

Furthermore, a second set of control variables was measured at different time points throughout each fMRI session and comprised six variables on subjective mood (alertness, contentedness, calmness, pleasure, arousal, and dominance) and four variables on physical wellbeing (pulse, systolic and diastolic blood pressure, and the side effects sum score). To test for drug effects on these variables, their scores at different time points after drug administration (t_1, t_2, t_3) were first subtracted by their baseline score before drug administration (t_0) for each drug condition. A repeated measures ANOVA with the factor drug was then performed on each of these difference scores ($t_1 - t_0$, $t_2 - t_0$, $t_3 - t_0$). The complete ANOVA results can be found in Table A3 of the appendix. To summarize, the ANOVA showed no significant drug effect for any of the physical wellbeing parameters (all $p > .05$). However, a significant drug effect ($F_{2,60} = 4.46$, $p = .016$) was found for one of the difference scores ($t_3 - t_0$) on the subscale "calmness" of the VAS (Bond & Lader, 1974). Paired t-tests on this variable revealed a significant increase (indicating lower calmness due to item reversal) under haloperidol compared to placebo (mean difference P-H = -0.67, $t_{30} = -2.05$, $p = .049$) and L-dopa (mean difference D-H = -0.85, $t_{30} = -2.99$, $p = .005$), but no significant difference between placebo and L-dopa (mean difference P-D = 0.18, $t_{30} = 0.61$, $p = .544$). It should be noted, however, that with the large number of hypothesis tests performed here (42 ANOVAs over all control variables and time points) and without correction for multiple comparisons, at least two significant results ($p < .05$) would be expected by chance alone, including one $p \leq .016$ (as it was found here) with a probability of ca. 50 %.

For completeness, also reaction times in the bandit task were tested for DA drug effects. However, a repeated measures ANOVA with the factor drug yielded no significant drug effect on either mean reaction time ($F_{2,60} = 0.54$, $p = .585$) or median reaction time ($F_{2,60} = 0..50$, $p = .611$).

Furthermore, to test whether subjects were actually blinded to the drug condition, their drug guesses after each session were examined for their correctness above chance. Over all subjects and drug sessions, 30 of the 93 guesses (32.3 %) were correct, which is in line with the number of correct guesses expected by chance (31, i.e. 33.3 %). However, this result does not rule out that some subjects performed above chance (i.e. recognized all drug conditions) and others below. Therefore, a second analysis was conducted based on the performance on the subject level, taking the within-subject design into account. First, the number of correct guesses per subject was calculated, resulting in the

following frequency distribution: 25.8 % (0 correct guesses), 51.6 % (1 correct guess), 22.6 % (2 correct guesses), 0.0 % (3 correct guesses). This distribution was then compared to the frequency distribution predicted by random guessing: 29.6 % (0 correct guesses), 44.4 % (1 correct guess), 22.2 % (2 correct guesses), 3.7 % (3 correct guesses). A chi-squared test showed that there was no significant difference between both distributions ($\chi^2 = 1.66$, $p = .634$; with Monte Carlo approximation). Finally, it was examined whether the frequency of drug guesses (i.e. how often each drug was guessed) depended on the actual drug condition. The corresponding data are reported in Table 8. A chi-squared test showed that the frequencies of drug guesses did not significantly differ between the three drug conditions ($\chi^2_4 = 0.36$, $p = .986$). Taken together, the results of all three analyses indicate that the observed data are in accordance with random guessing.

Table 8. Frequencies of drug guesses for each drug condition.

drug condition	guess: placebo	guess: L-dopa	guess: haloperidol
placebo (n=31)	18	5	8
L-dopa (n=31)	17	5	9
haloperidol (n=31)	19	5	7

Finally, subjects' confidence ratings for drug guesses were analyzed. Over all subjects and drug conditions, the average confidence rating was 2.65 ($SD = 1.03$), i.e. between 2 ("uncertain") and 3 ("moderately certain"). Confidence ratings did not significantly differ between drug conditions (ANOVA: $F_{2,56} = 0.51$, $p = .605$) or between correct and incorrect guesses (two-sample t-test: $t_{60.27} = -0.10$, $p = .922$).

5.5 Inverted-U analysis

5.5.1 Modulation of choice behavior by the individual DA baseline

The inverted-U hypothesis of DA states that DA-dependent cognitive functions are modulated by the individual DA baseline according to an optimum (inverted-U-shaped) curve (see 1.2.4). As the trade-off between exploration and exploitation is assumed to be a DA-dependent behavior, it should be modulated by the individual DA baseline accordingly. To test for this assumption, it was examined whether individual differences in explore/exploit behavior, as assessed by different model-based and model-free choice variables, were predicted by the individual DA baseline (indexed by the sEBR and WMC) according to an inverse quadratic relationship. To increase sample size for this analysis, data from the placebo condition (n=31) and pilot study 2 (n=16) were combined, resulting in a sample of 47 subjects.

In preparation for this analysis, a principal component analysis (PCA) was performed on the z-transformed WMC data, including task scores from the Rotation Span Task, Operation Span Task, and

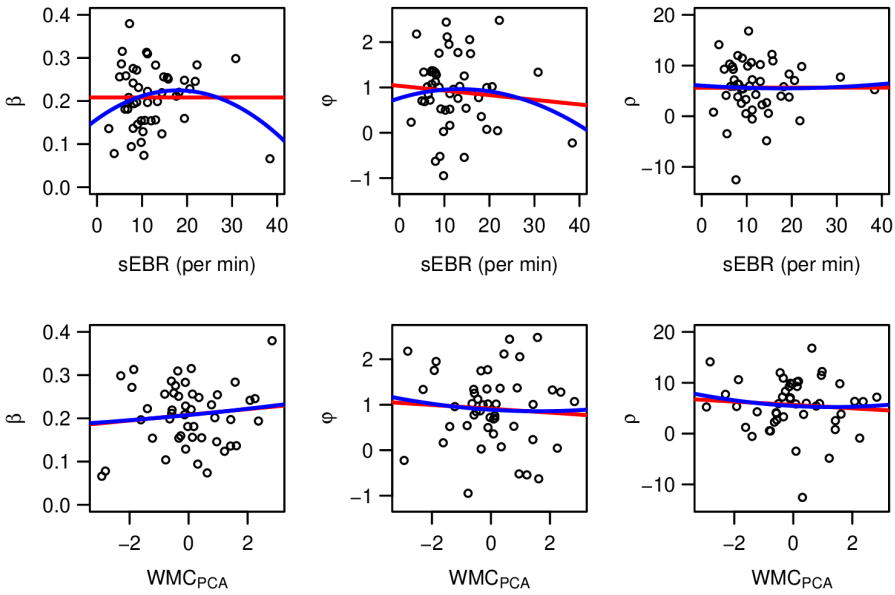
Listening Span Task. Loadings on the first principal component, which explained 56.6 % of the combined variance, were 0.62 for the Rotation Span Task, 0.64 for the Operation Span Task, and 0.46 for the Listening Span Task. Subjects' scores on the first principal component were then used as a WMC compound measure, denoted as WMC_{PCA}, for the following plots and analyses.

Results are first reported for the model-based choice variables, including the three choice parameters (β, φ, ρ) of the Bayes-SM+EP model. For each of these parameters, subject-level posterior medians were plotted against the two baseline DA proxies sEBR and WMC_{PCA} (see Figure 15a). Two regression models were then fitted to the data shown in each plot: a "linear model" (LM, red line) testing only for a linear relationship between a given choice variable and DA proxy, and a "quadratic model" (QM, blue line) testing also for a quadratic relationship between these variables (see 2.9.2). Next, both models were compared using leave-one-out (LOO) cross-validation. Therefore, LOO measures based on squared distances were calculated for each model and subtracted (LOO_{LM} - LOO_{QM}), such that negative values indicate a higher predictive accuracy of the LM. The results of this LOO model comparison are reported in the upper part of Table 9, showing only negative values for all six comparisons, i.e. higher predictive accuracy of the LM. Additionally, the estimate and p-value for the β_2 parameter of each QM are reported in Table 9 (upper part), showing that the β_2 parameter does not significantly differ from zero for any of the six QMs.

Next, the same kind of analysis was conducted for the four model-free choice variables, i.e. *payout*, *% best bandit*, *mean rank,* and *% switches*. First, each of these variables was plotted against the sEBR and WMC_{PCA} (see Figure 15b), then both an LM and a QM were fitted to the data shown in each plot. Results of the LOO model comparison for the model-free variables are presented in the lower part of Table 9, along with estimates and p-values for the β_2 parameters of all QMs. Most of these LOO comparisons yielded negative values, indicating a higher predictive accuracy of the LM. Only the regressions of *% best bandit* and *mean rank* on the sEBR showed a higher predictive accuracy of the QM. Yet, a visual inspection of the two respective plots in Figure 15b shows that the right part of the QM fit is only informed by very few data due to the asymmetric distribution of sEBR values in the sample. Also, none of the eight QMs showed a β_2 estimate that significantly differed from zero.

Finally, the same kind of analysis was repeated for all dependent variables (model-based and model-free choice measures) using the separate WMC tasks scores (Rotation Span, Operation Span, or Listening Span) as baseline DA proxies instead of the WMC_{PCA} score. All these LOO comparisons showed a higher predictive accuracy of the LM compared to the QM, and there was no QM for which β_2 estimates significantly differed from zero.

a model-based choice variables



b model-free choice variables

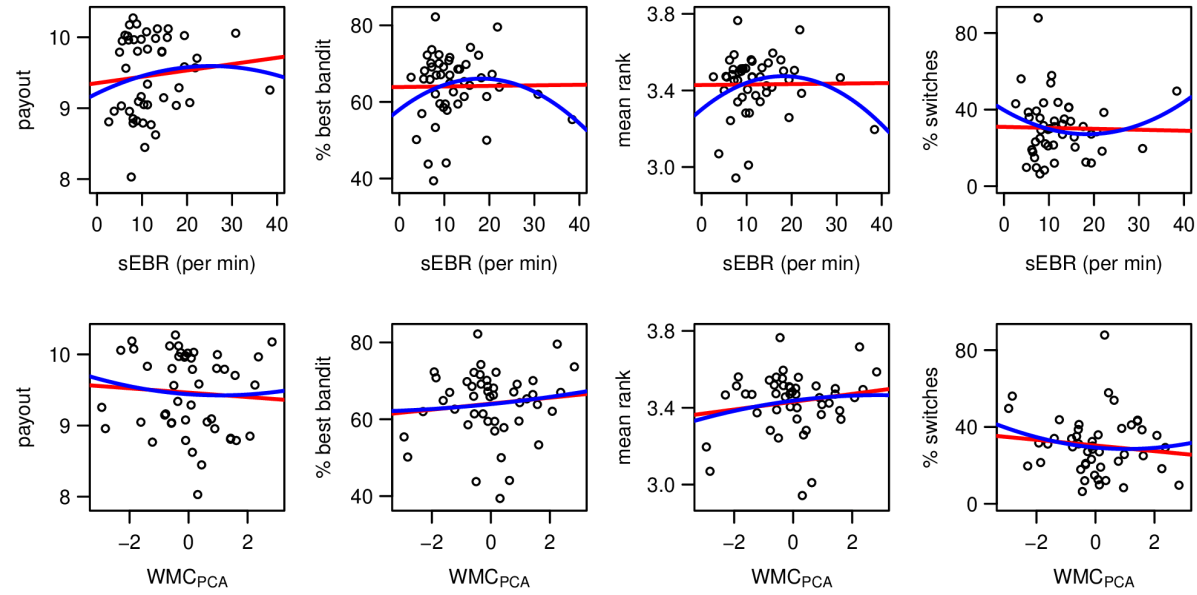


Figure 15. Test for an inverted-U relationship between choice behavior and DA baseline. Choice behavior was assessed by (**a**) the posterior medians of the three choice parameters (β, φ, ρ) of the Bayes-SM+EP model and (**b**) the four model-free choice variables (*payout*, *% best bandit, mean rank, % switches*). Baseline dopamine (DA) function was assessed by the two behavioral DA proxies spontaneous eye blink rate (sEBR) and working memory capacity (WMC). For the latter, the first principal component across three different WMC tasks was used, denoted by WMC_{PCA}. Each plot shows two regression lines that were fitted to the data, one for the linear model (red line) and one for the quadratic model (blue line). Note that data from pilot study 2 and the placebo condition of the main study were combined for this analysis to increase the sample size to n=47. β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter.

Table 9. Test for an inverted-U relationship between choice behavior and DA baseline.

	LOO_{LM} - LOO_{QM}		β_2 estimate		β_2 *p*-value	
	sEBR	WMC_{PCA}	sEBR	WMC_{PCA}	sEBR	WMC_{PCA}
model-based:						
β	-0.06	-0.04	$-2.09e^{-04}$	$2.98e^{-04}$	.132	.949
φ	-3.60	-2.57	$-1.13e^{-03}$	$1.27e^{-02}$	.470	.809
ρ	-53.09	-49.68	$1.69e^{-03}$	$1.20e^{-01}$	.869	.726
model-free:						
payout	-0.95	-1.05	$-6.04e^{-04}$	$1.37e^{-02}$	.582	.710
% best bandit	198.06	-245.78	$-2.45e^{-02}$	$7.40e^{-02}$	.149	.897
mean rank	0.06	-0.10	$-5.29e^{-04}$	$-3.45e^{-03}$	.080	.733
% switches	-484.04	-700.67	$-2.09e^{-04}$	$6.58e^{-01}$	.222	.509

Note. Choice behavior was assessed by the three choice parameters of the Bayes-SM+EP model (upper part) and four model-free choice variables (lower part). Baseline dopamine (DA) function was assessed by the two behavioral DA proxies spontaneous eye blink rate (sEBR) and working memory capacity (WMC). For the latter, the first principal component across three different WMC tasks was used, denoted by WMC_{PCA}. The column "LOO_{LM} - LOO_{QM}" shows the difference of the squared distances for the linear model (LM) minus the quadratic model (QM) from the leave-one-out (LOO) model comparison. Note that negative values for LOO_{LM} - LOO_{QM} indicate better predictive accuracy of the LM. The columns "β_2 estimate" and "β_2 *p*-value" show for each quadratic model the estimated value and *p*-value of the β_2 regression coefficient, respectively. Note that data from pilot study 2 and the placebo condition of the main study were combined for this analysis to increase the sample size to n=47. β*:* softmax parameter; φ*:* exploration bonus parameter; ρ*:* perseveration bonus parameter.

5.5.2 Modulation of behavioral drug effects by the individual DA baseline

According to the inverted-U hypothesis of DA, the direction and magnitude of DA drug effects depend on the individual DA baseline (see 1.2.4). To test for this hypothesis, it was examined if L-dopa and haloperidol effects on explore/exploit behavior were modulated by the individual DA baseline, as indexed by the sEBR and WMC.

Results for the exploration bonus parameter φ are reported first to demonstrate the different steps of this analysis. First, the subject-level φ medians were subtracted pairwise between all drug conditions (P-D, P-H, D-H) to quantify the direction and magnitude of DA drug effects for each subject. These φ drug differences were then plotted against the two baseline DA proxies sEBR and WMC_{PCA} (see Figure 16). Visual inspection of these plots did not show any systematic variation of the DA drug effects (y-axes) dependent on the DA baseline measures (x-axes), contrary to the assumption of the inverted-U hypothesis. To further test this, subjects were divided by median split into a low (n=15) and high (n=16) DA baseline group, once for the sEBR and once for the WMC_{PCA}. Then, two-sample t-tests were performed on these data to compare the φ drug differences between the low and the high DA baseline group. The t-tests showed no significant difference between the low and the high DA baseline group for any of the six comparisons (see Table 10).

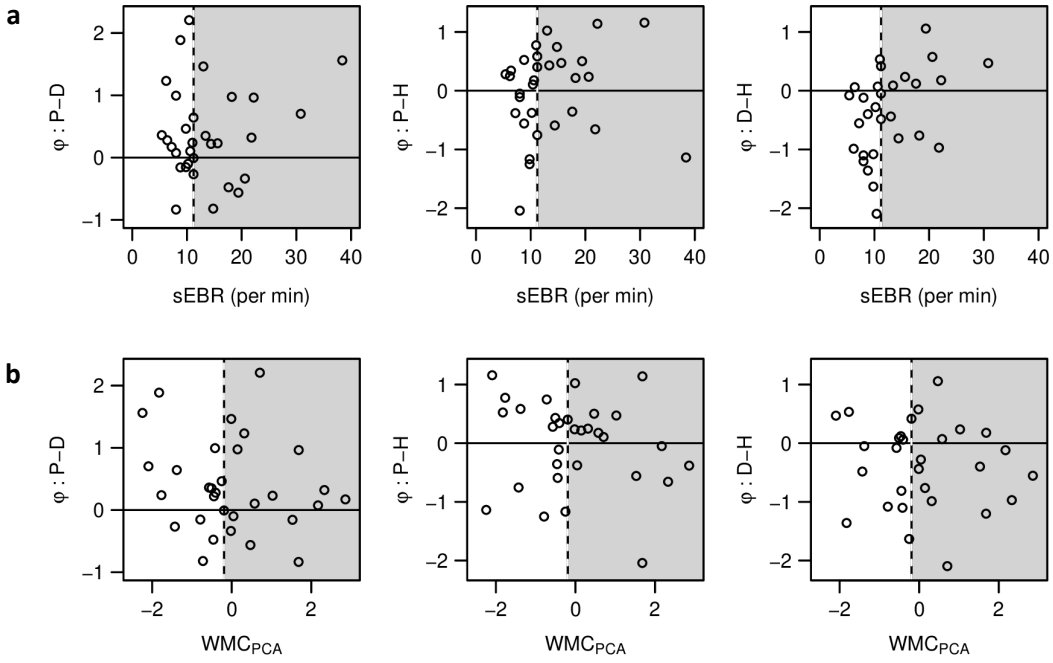


Figure 16. Drug effects on directed exploration in dependence of the DA baseline. Baseline dopamine (DA) function was assessed by the two behavioral DA proxies (**a**) spontaneous eye blink rate (sEBR) and (**b**) working memory capacity (WMC). For the latter, the first principal component across three different WMC tasks was used, denoted by WMC_{PCA}. Drug effects on directed exploration were assessed by calculating pairwise drug differences of the subject-level exploration bonus parameter (φ) medians. Each plot is split at the median of the DA proxy (sEBR or WMC_{PCA}) into a white area (below the median) and a grey area (above the median). P: placebo; D: L-dopa; H: haloperidol.

Table 10. Comparison of drug effects on choice behavior between low and high DA baseline.

	sEBR			WMC_{PCA}		
model parameter	low	high	group comparison	low	high	group comparison
β						
P - D	0.012	0.010	$t_{27.77}$ = 0.07; p = .943	0.003	0.019	$t_{26.93}$ = -0.57; p = .576
P - H	-0.019	0.032	$t_{28.71}$ = -1.94; p = .062	-0.011	0.024	$t_{28.65}$ = -1.32; p = .199
D - H	-0.031	0.021	$t_{26.60}$ = -1.94; p = .063	-0.014	0.006	$t_{24.63}$ = -0.70; p = .492
φ						
P - D	0.451	0.310	$t_{27.92}$ = 0.51; p = .611	0.399	0.359	$t_{28.92}$ = 0.15; p = .885
P - H	-0.232	0.213	$t_{28.49}$ = -1.69; p = .103	-0.035	0.029	$t_{28.57}$ = -0.23; p = .819
D - H	-0.682	-0.095	$t_{27.81}$ = -1.90; p = .068	-0.431	-0.330	$t_{25.80}$ = -0.31; p = .762
ρ						
P - D	2.085	-0.765	$t_{28.51}$ = 1.77; p = .088	0.438	0.779	$t_{28.95}$ = -0.20; p = .842
P - H	-0.077	-1.522	$t_{28.92}$ = 1.33; p = .195	-0.752	-0.889	$t_{24.14}$ = 0.12; p = .906
D - H	-2.173	-0.751	$t_{27.31}$ = -0.84; p = .408	-1.192	-1.670	$t_{25.22}$ = 0.28; p = .783

Note. Group comparisons were performed by two-sample t-tests, using the pairwise drug differences of the subject-level choice parameter medians of the Bayes-SM+EP model as dependent variables. Baseline dopamine (DA) function was assessed by the two behavioral DA proxies spontaneous eye blink rate (sEBR) and working memory capacity (WMC). For the latter, the first principal component across three different WMC tasks was used, denoted by WMC_{PCA}. β: softmax parameter; φ: exploration bonus parameter; ρ: perseveration bonus parameter; P: placebo; D: L-dopa; H: haloperidol.

In addition, the same kind of analysis was performed with the two remaining choice parameters (β, ρ) as dependent variables (see Table 10), and also for all three dependent variables using the separate WMC task scores (Rotation Span, Operation Span, or Listening Span) as DA proxies instead of the WMC_{PCA} score. Altogether, no evidence was found that DA drug effects on any of these choice measures were modulated by the individual DA baseline according to an inverted-U function, neither from visual inspection of the plotted data (not shown), nor from the two-sample t-tests comparing DA drug effects between the low and the high DA baseline group (all $p > .05$).

5.6 fMRI results

5.6.1 Brain activity associated with exploration and exploitation

First, differences in brain activity between exploratory and exploitative choices were analyzed across all subjects and drug conditions. For this, trials were classified as either exploitative (following the highest expected value) or exploratory (not following the highest expected value) as previously described by Daw et al. (2006). It was found that the pattern of brain activity markedly differed between both types of choices, as shown in Figure 17.

Exploratory choices were associated with greater activation in the bilateral frontopolar cortex (FPC; left: -42, 27, 27 mm; $z = 6.07$; right: 39, 34, 28 mm; $z = 7.56$) and in a large cluster along the bilateral intraparietal sulcus (IPS; cluster peak at -48, -33, 52; $z = 10.45$), extending on both sides into the postcentral gyrus and precuneus. Furthermore, greater activation during exploratory trials was also observed bilaterally in the anterior insula (AI; left: -36, 15, 3 mm; $z = 6.69$; right: 36, 20, 3 mm; $z = 6.87$) and in a cluster extending into the dorsal anterior cingulate cortex (dACC; cluster peak at 8, 12, 45 mm; $z = 8.47$). Also, the thalamus, cerebellum, and supplementary motor area showed increased bilateral activation during exploration compared to exploitation. A complete list of activations associated with exploratory choices can be found in Table A6 of the appendix.

In contrast, exploitative choices were associated with greater activation in the ventromedial prefrontal cortex (vmPFC; -2, 40, -10 mm; $z = 5.67$) and bilaterally in the lateral orbitofrontal cortex (lOFC; left: -38, 34, -14 mm; $z = 5.81$; right: 38, 36, -12 mm; $z = 5.02$). Furthermore, greater activation during exploitative trials was also observed in a cluster spanning the left posterior cingulate cortex (PCC) and left precuneus (cluster peak at -6, -52, 15 mm; $z = 7.40$), as well as bilaterally in the angular gyrus (left: -42, -74, 34 mm; $z = 8.04$; right: 52, -68, 28 mm; $z = 7.02$), hippocampus (left: -24, -16, -15 mm; $z = 4.16$; only at $p < .001$, uncorrected; right: 32, -16, -15 mm; $z = 5.09$), and several clusters along the superior and middle temporal gyrus. A complete list of activations associated with exploitative choices can be found in Table A7 of the appendix. Aside from these neural correlates of exploratory and exploitative choices, it was found that the reward prediction error positively correlated with activity in the bilateral ventral striatum (left: -16, 6, -12 mm; $z = 6.40$; right: 16, 9, 10 mm; $z = 6.20$), as shown in Figure 18.

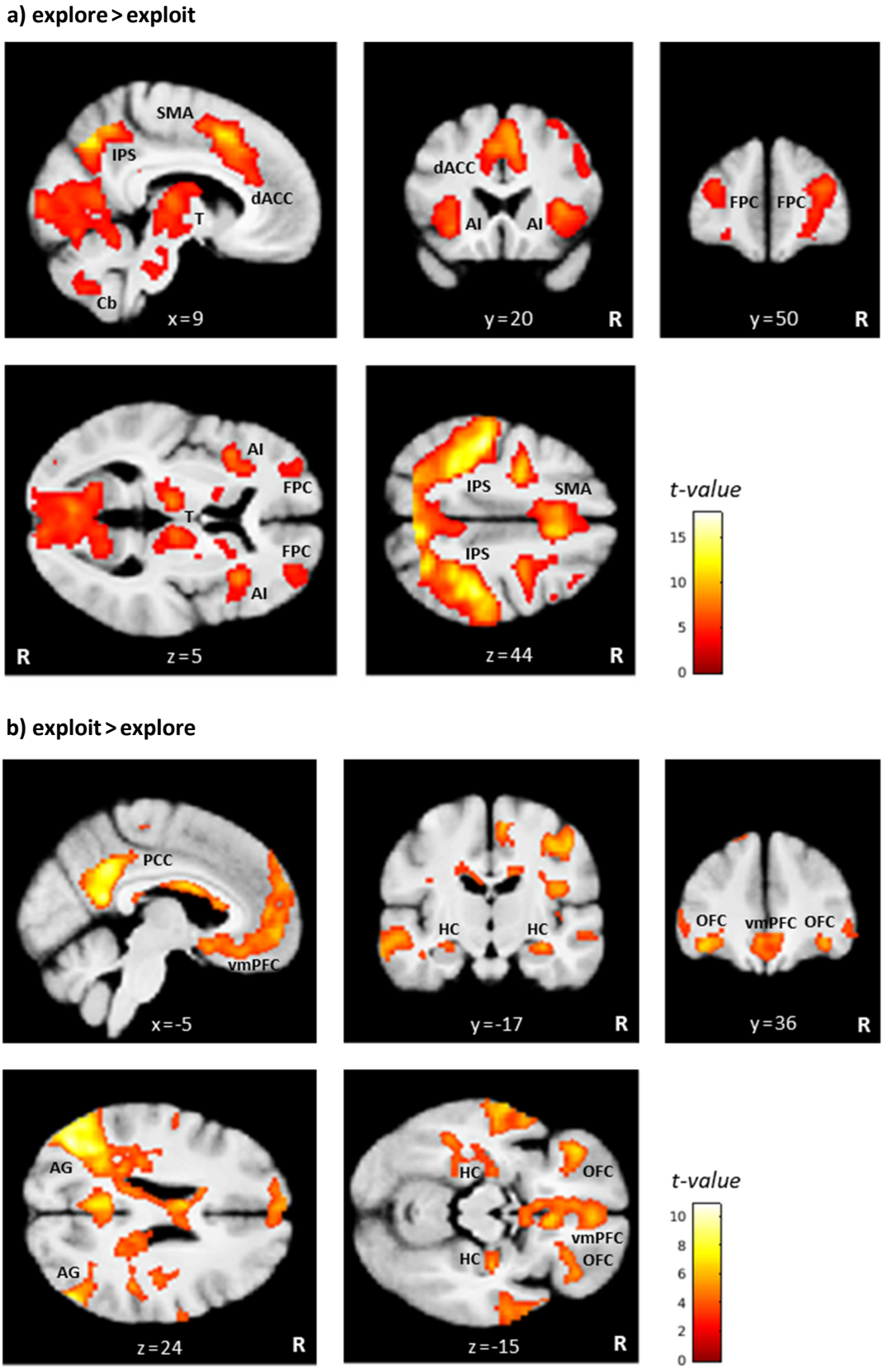


Figure 17. Brain regions differentially activated by exploratory and exploitative choices. Shown are statistical parametric maps for (**a**) the contrast *explore > exploit* and (**b**) the contrast *exploit > explore* over all drug conditions. AG: angular gyrus; AI: anterior insula; Cb: cerebellum; dACC: dorsal anterior cingulate cortex; FPC: frontopolar cortex; HC: hippocampus; IPS: intraparietal sulcus; vmPFC: ventromedial prefrontal cortex; OFC: orbitofrontal cortex; PCC: posterior cingulate cortex; SMA: supplementary motor area; T: thalamus. Thresholded at $p < .001$, uncorrected. R: right.

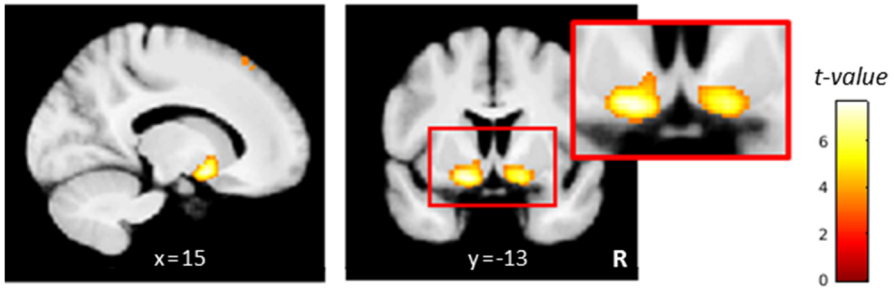


Figure 18. Striatal coding of the reward prediction error**.** Activity in the bilateral ventral striatum correlated positively with the reward prediction error signal. Thresholded at $p < .001$, uncorrected. R: right.

Next, it was analyzed whether the subdivision of exploratory choices into directed explorations (following the highest exploration bonus) and random explorations (not following the highest exploration bonus) revealed different activation patterns for both types of explorations. Therefore, a second GLM was set up that included two separate regressors for directed and random explorations (both modeled at trial onset). While the contrast *directed > random* yielded no suprathreshold activations across drugs or for any of the drug conditions alone, the opposite contrast (*random > directed*) yielded a small cluster of three voxels in the right frontopolar cortex (32, 50, -8 mm; $z = 5.34$) across drugs, though not for any of the drug conditions alone. It should be noted, however, that the number of trials was very unequal for both exploration conditions with considerably more trials in the random exploration condition (on average 3.2 times more random than directed exploration trials per session). Hence, the second-level results for this contrast should be treated with caution (see Chen, Saad, Nath, Beauchamp, & Cox, 2012). Also, after exclusion of all sessions with ≤ 5 trials in the directed exploration condition (excluding 8 out of 93 sessions), the frontopolar cluster for the contrast *random > directed* was no longer significant. Furthermore, overlaying activation maps for directed, random, and overall explorations (each contrasted against exploitation) showed substantially the same activation pattern under all three exploration conditions (see Figure 19), including the bilateral FPC, IPS, dACC, AI, and thalamus. Note that for the overlay, activation maps were displayed at a liberal threshold of $p < .05$ (uncorrected) to account for the small trial number and low statistical power in the directed exploration condition.

In addition to the main GLM with a binary coding of explore/exploit (1/0), a third GLM was set up to examine instead the parametric effects of two model-based quantities that are tightly involved in exploitation and exploration: the expected value ($\hat{\mu}^{pre}$) and uncertainty ($\hat{\sigma}^{pre}$) of the chosen bandit (both modeled at trial onset). While the expected value was positively correlated with activity in a network of brain regions largely overlapping with the one for exploitative choices (see Figure 20a), the uncertainty was positively correlated with activity in a network of brain regions largely overlapping with the one for exploratory choices (see Figure 20b). Although results of this third GLM are not reported in further detail, it is important to note that the expected value and uncertainty are both quantities that are substantially correlated with the choice type explore/exploit (see Table A4 in the

appendix), which should be kept in mind when interpreting the fMRI results of the main GLM (as discussed in section 6.3.1).

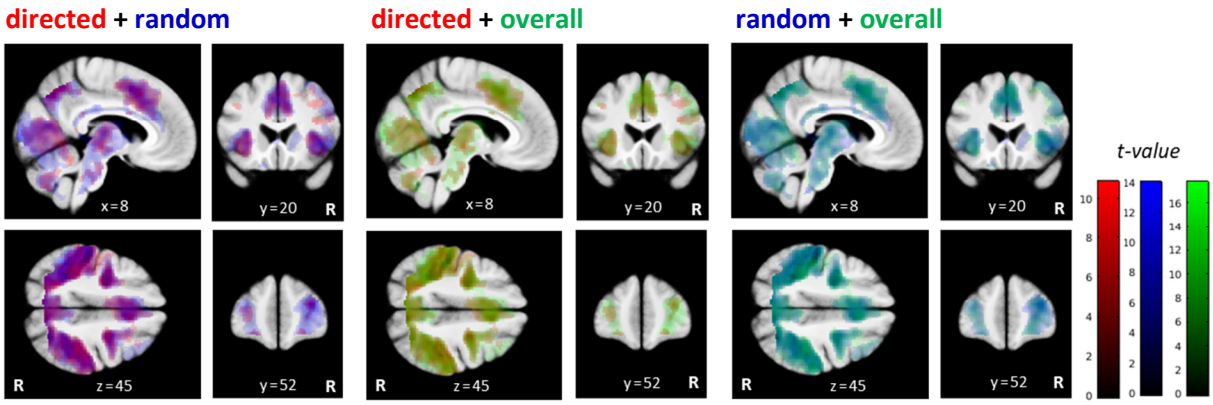


Figure 19. Brain activation patterns for different types of exploration**.** Shown are pairwise overlays of the statistical parametric maps for the contrasts *explore > exploit* ("overall" in green), *directed > exploit* ("directed" in red), and *random > exploit* ("random" in blue) over all drug conditions. While the first contrast is based on a binary choice classification according to which all choices not following the highest expected value are explorations, the other two contrasts are based on a trinary choice classification, which further subdivides explorations into choices following the highest exploration bonus (directed) and choices not following the highest exploration bonus (random). Thresholded at $p < .05$, uncorrected for display purposes. R: right.

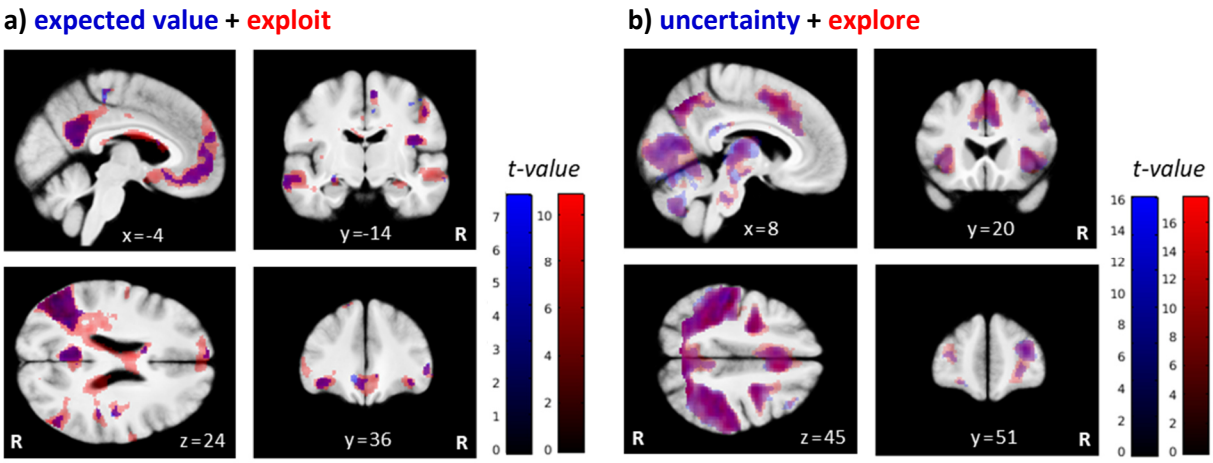


Figure 20. Neural codings of expected value and uncertainty. (**a**) Overlay of the statistical parametric maps for the parametric regressor *expected value* (in blue) and the contrast *exploit > explore* ("exploit" in red) over all drug conditions. (**b**) Overlay of the statistical parametric maps for the parametric regressor *uncertainty* (in blue) and the contrast *explore > exploit* ("explore" in red) over all drug conditions. Thresholded at $p < .001$, uncorrected. R: right.

5.6.2 Brain activity differences between drug conditions (planned comparisons)

First, it was analyzed whether brain activation patterns for exploratory and exploitative choices showed any differences between the DA drug conditions across subjects. Therefore, a repeated measures ANOVA was performed on the second level to test for the main effect of drug on the contrast *explore vs. exploit* of the first GLM, and also on the contrasts *directed vs. exploit*, *random vs. exploit,*

and *random vs. directed* of the second GLM. However, none of these tests yielded any suprathreshold activations on the whole-brain level, nor in any of the seven regions included in the small volume correction (i.e. left/right FPC, left/right IPS, left/right AI, and dACC). In addition, the same kind of analysis was performed on the four remaining regressors of the first GLM (*trial onset, reward onset, prediction error,* and *outcome value*), also revealing no suprathreshold activations for any of these regressors on the whole-brain level.

According to the behavioral results, DA drug effects on directed exploration (i.e. the φ parameter) largely differed between subjects in magnitude and direction (see Figure 13). Therefore, a second-level regression analysis was performed for each drug pair, testing whether drug effects on exploration-specific brain activity (i.e. pairwise drug differences of the subject-specific contrast images *explore vs. exploit, directed vs. exploit,* and *random vs. exploit*) were linearly predicted by the drug effects on exploratory behavior (i.e. pairwise drug differences of the subject-specific φ medians). However, none of these regression analyses revealed any suprathreshold activations on the whole-brain level, nor in any of the seven regions included in the small volume correction (see above).

5.6.3 Exploratory fMRI analysis

Following these planned analyses, fMRI data were further explored for DA drug effects on the neural level that might explain the behavioral findings, foremost the reduction of directed exploration under L-dopa compared to placebo. Based on the null findings in the planned analyses, it was reasoned that L-dopa might exhibit its influence on the explore/exploit trade-off not by altering brain activations for exploratory or exploitative choices per se, but rather by affecting the neural correlates involved in behavioral switching from exploitation to directed exploration when the overall uncertainty increases. Thereby, L-dopa might delay the time point at which directed exploration is triggered in response to accumulating uncertainty, hence resulting in fewer directed explorations over time. This alternative hypothesis was tested by performing another model-based fMRI analysis. According to the cognitive model (Bayes-SM+EP), one important quantity to trigger directed exploration at a specific time point is the growing uncertainty over all choice options that are currently not exploited, since this is directly linked (via the exploration bonus) to the probability to choose (explore) one of these options. This overall uncertainty ($\Sigma\hat{\sigma}^{pre}$) can be quantified in each trial by the summed standard deviation ($\hat{\sigma}^{pre}$) of all four bandits, whereby $\hat{\sigma}^{pre}$ of the currently exploited bandit is constant over trials and negligible in this process. The resulting metric gradually increases during a series of exploitations, but reduces abruptly when one or more bandits with high uncertainty are explored (see Figure 9f). Following this reasoning, the overall uncertainty was used as a parametric regressor in a new GLM (modeled at trial onset) to reveal brain regions for which activity was correlated with this quantity. The contrast images for this regressor were then used in a second-level random effects analysis with the factors drug and subject. Since a comparison between the placebo and the L-dopa condition was of primary interest in this exploratory fMRI analysis (based on the behavioral findings), these two drug conditions were

analyzed first. In the placebo condition alone, no voxels survived FWE correction ($p < .05$), but a more lenient threshold ($p < .001$, uncorrected) showed that activity in the bilateral dACC (cluster peak at -3, 21, 39 mm; $z = 3.96$), right anterior insula (42, 15, -6 mm; $z = 3.46$), and left posterior insula (-34, -20, 8 mm; $z = 4.63$) was positively correlated with the overall uncertainty (see Figure 21a). Next, the placebo and L-dopa condition were compared by directed t-contrasts (*placebo > L-dopa* and *L-dopa > placebo*) to find regions for which the parametric effect of the overall uncertainty differed between both drug conditions. While the contrast *L-dopa > placebo* yielded no suprathreshold activations, the opposite contrast (*placebo > L-dopa*) revealed a significant activation in the left posterior insula (-34, -20, 8 mm; $z = 5.05$). At a reduced threshold ($p < .001$, uncorrected), also activity in the left anterior insula (-38, 6, 14 mm; $z = 4.88$) and bilateral dACC (left: -2, 36, 33 mm; $z = 3.32$; right: 4, 14, 28 mm; $z = 3.41$) showed a stronger correlation with the overall uncertainty under placebo compared to L-dopa (see Figure 21b). For completeness, the placebo and L-dopa condition were also compared to haloperidol by computing four directed t-contrasts (*placebo > haloperidol*, *haloperidol > placebo*, *L-dopa > haloperidol*, *haloperidol > L-dopa*), none of which yielded any suprathreshold activations. However, at a lower threshold ($p < .001$, uncorrected), the contrast *placebo > haloperidol* revealed activations in a number of regions, including the bilateral anterior insula (left: -30, 21, 6 mm; $z = 3.81$; right: 39, 15, -4 mm; $z = 3.88$) and left posterior insula (-34, -22, 8 mm; $z = 3.78$), as shown in Figure 21c. A complete list of activations for all t-contrasts of this second-level analysis can be found in Table A8 of the appendix.

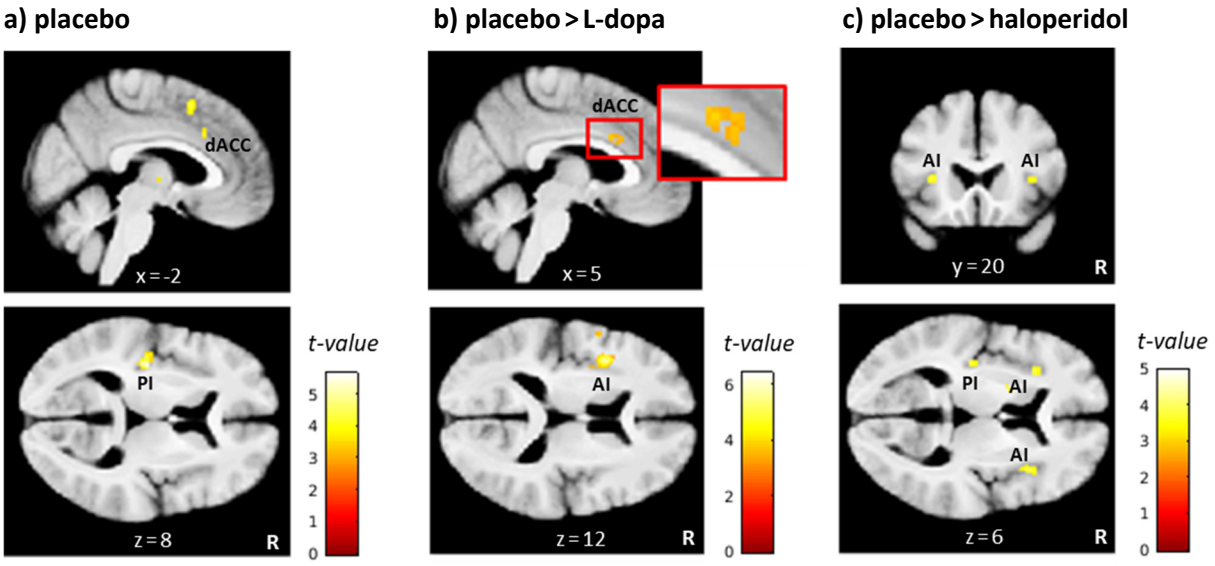


Figure 21. Drug effects on the neural codings of overall uncertainty. **(a)** Regions in which activity correlated positively with the overall uncertainty in the placebo condition included the dorsal anterior cingulate cortex (dACC) and left posterior insula (PI). (**b**) Regions in which the correlation with overall uncertainty was reduced under L-dopa compared to placebo included the dACC and left anterior insula (AI). (**c**) Regions in which the correlation with overall uncertainty was reduced under haloperidol compared to placebo included the bilateral AI and left PI. Thresholded at $p < .001$, uncorrected. R: right.

Finally, a second-level regression analysis was performed on the *overall uncertainty* regressor for each drug pair. This regression analysis tested whether drug effects on neural activations for the overall uncertainty (i.e. pairwise drug differences of the subject-specific contrast images for *overall uncertainty*) were linearly predicted by the drug effects on directed exploration behavior (i.e. pairwise drug differences of the subject-specific φ medians). However, this regression analysis yielded no suprathreshold activations for any of the three pairwise drug comparisons.

6 Discussion

6.1 Summary of results

The current study examined the causal role of DA in human explore/exploit behavior in a pharmacological fMRI approach, using L-dopa (DA precursor) and haloperidol (DA antagonist) in a double-blind, placebo-controlled, counterbalanced, within-subjects design. First, explore/exploit behavior, as assessed with the restless four-armed bandit task, was analyzed using different cognitive models of learning and decision making in a hierarchical Bayesian modeling approach. A quantitative model comparison showed that choice behavior was best described by the Bayes-SM+EP model, which combines a Bayesian learning rule tracking both mean and variance (uncertainty) of the expected reward, with a modified softmax choice rule capturing both random and directed exploration along with choice perseveration. Using this model, it was found that directed (uncertainty-driven) exploration, as indexed by the φ parameter, was significantly reduced across subjects under L-dopa compared to placebo. In contrast, haloperidol did not significantly shift the φ parameter across subjects, but showed a tendency to reduce the group-level variance of this parameter (Λ^{φ}) relative to placebo and L-dopa. No overall drug effects were observed on random exploration or perseveration, as respectively indexed by the β and ρ parameter, nor on any of the tested model-free choice variables. Also, neither drug was found to cause apparent side effects on physical well-being, self-reported mood, or alertness, hence these factors may be ruled out as potential mediators of the observed drug effects. To examine drug effects on the neural level, choices were first classified as either exploitative (i.e. following the highest expected reward value) or exploratory, and the pattern of brain activity was compared between both types of choices. Across all drug conditions, exploratory choices were associated with higher activity in the FPC, IPS, dACC, and insula, whereas exploitative choices showed higher activity in the vmPFC and OFC, as well as in the PCC, precuneus, angular gyrus, and hippocampus, largely replicating the results of previous studies (see 1.1.4). Surprisingly, no drug effects were found on these neural correlates of exploratory and exploitative choices, nor on striatal reward prediction error signaling. Yet, an exploratory analysis of the brain imaging data revealed that L-dopa reduced insular and dACC activity associated with the overall reward uncertainty across all choice options. Finally, no evidence was found in support of the added hypothesis that DA drug effects on explore/exploit behavior are modulated in an inverted-U fashion by the individual DA baseline, as assessed by the behavioral proxies sEBR and WMC.

6.2 Behavioral results

Overall, the finding that pharmacological manipulation of the DA system resulted in a shift of the explore/exploit trade-off is in line with previous animal and human studies suggesting that DA is causally involved in regulating this trade-off (see 1.3). However, the observed pattern of DA drug effects on explore/exploit behavior did not match the initial hypothesis, according to which both

random and directed exploration should be increased under L-dopa vs. placebo and decreased under haloperidol vs. placebo. In the following, the behavioral results of the L-dopa treatment will be discussed first, followed by a discussion of the haloperidol results.

6.2.1 L-dopa effects

Model-based analysis of choice behavior revealed that across subjects, L-dopa administration resulted in a reduction of the φ parameter (capturing directed exploration) compared to placebo, while the β parameter (capturing random exploration) remained unaffected by the drug. It should be noted first that this reduction of the φ parameter can be regarded from two perspectives: On the one hand, it indicates a reduced tendency for uncertainty-driven exploration, but on the other hand also an increased tendency for value-driven exploitation. Remember that the φ parameter of the cognitive model determines the relative degree to which choices are biased towards the uncertainty of an option, whereby a smaller absolute value of φ indicates that less weight is given to the uncertainty and, in turn, more weight to the value (which can be thought of as having a constant weight of one in this model). Hence, the observed choice behavior under L-dopa can be described as less uncertainty-driven and more value-driven. Accordingly, when classifying all choices per subject into exploitations, directed explorations, and random explorations, L-dopa was found to reduce the percentage of directed (but not random) explorations compared to placebo across subjects, and to marginally increase the percentage of exploitations. Since on the neural level, L-dopa is well known to increase DA transmission compared to placebo, these findings suggest that increased DA strengthens value-driven, exploitative behavior against uncertainty-driven exploration. While this interpretation is in line with several studies showing that *striatal* DA drives reinforcement learning and exploitative behavior (e.g. Frank et al., 2004, 2009; Moustafa et al., 2008; Pessiglione et al., 2006), it contradicts other studies suggesting that *prefrontal* DA promotes uncertainty-driven exploration (e.g. Blanco et al., 2015; Frank et al., 2009; Kayser et al., 2015; see 1.3.2). Thus, to compare the L-dopa results with other studies on DA's role in explore/exploit behavior, it is important to distinguish striatal from prefrontal DA effects and to first regard these two subsystems separately. Therefore, the L-dopa findings are first discussed with respect to striatal DA function and then with respect to prefrontal DA.

6.2.1.1 L-dopa effects with respect to striatal DA function

To begin with, previous studies support the view that L-dopa promotes positive reinforcement and exploitation by enhancing striatal DA activity. For example, Pessiglione et al. (2006) showed in a reinforcement learning task with monetary gains and losses that administration of L-dopa compared to haloperidol biases choice behavior towards the most rewarding action (i.e. increases exploitation) and also increases the magnitude of the reward prediction error signal in the striatum. Specifically, this drug effect was only observed under the task condition of positive reinforcement ("gain condition"), but not negative reinforcement ("loss condition"). Moreover, the study demonstrated that the drug-

induced differences in striatal activity could explain the drug effects on choice behavior when incorporated into a standard reinforcement learning model. To show this, they first used the amplitude of the striatal reward prediction error signal to estimate the effective reward value for each drug condition, which was £1.29 for the L-dopa group and £0.71 for the haloperidol group (compared to a reference value set to £1.0 for the placebo control group). These drug-specific reward values were then incorporated into the reinforcement learning model, showing that they accurately and specifically reproduced the observed drug effects on choice behavior. Together, these results provide strong evidence for a causal involvement of striatal DA in driving exploitative behavior, whereby increased DA by L-dopa seems to promote more value-driven choices by increasing the apparent value of rewards as represented in the striatum. Integrating these results with the findings of the current study, it might be hypothesized that L-dopa administration increased the apparent reward value compared to placebo and haloperidol, such that obtained rewards were coded as "more valuable" in the striatum, which in turn might explain the shift towards more value-driven, exploitative behavior under L-dopa (but see 6.3 for fMRI results).

Furthermore, the effects of L-dopa administration on reward-based decision making were often examined in Parkinson's disease (PD) by comparing patients on and off dopaminergic medication (Frank et al., 2004; Moustafa et al., 2008; Rutledge et al., 2009). These studies consistently found that PD patients on DA medication show enhanced positive reinforcement ("Go learning") compared to patients off medication, whereas negative reinforcement ("NoGo learning") is either not affected or impaired by DA medication. For example, Moustafa et al. (2008) studied the choice behavior of PD patients in the clock task (see 1.1.2) and found that patients on medication were better at speeding up decisions to maximize expected rewards compared to the same patients off medication. Similar enhancements of Go learning in PD patients on versus off medication were also found in reinforcement learning tasks with either probabilistic or deterministic outcomes (Frank et al., 2004) and in a dynamic foraging task (Rutledge et al., 2009). Hence, these studies support the view that DA plays an important role in learning from positive outcomes and therefore also in guiding exploitative behavior. Importantly, all three studies attributed the DA medication effects on reinforcement learning specifically to *striatal* DA function. While none of these studies directly provided neuroimaging data, it still appears reasonable to interpret the above findings in terms of striatal DA function for different reasons. First, L-dopa is known to primarily act on the level of the striatum, while having much smaller effects on the prefrontal DA system, which is depleted to a lesser degree in early PD (see reviews by Cools, 2006; Lloyd et al., 1975). This (relative) regional specificity of L-dopa effects might potentially be explained by the finding that striatal regions exhibit the highest DOPA decarboxylase activity, the enzyme that converts L-dopa to DA in the brain (see Hälbig & Koller, 2007; Hefti & Melamed, 1980; Lloyd et al., 1975; Lloyd & Hornykiewicz, 1972; Melamed et al., 1980). Indeed, neurochemical studies in rats suggest that L-dopa administration generates 50-60 times more DA in the intact striatum than in the cortex (Carey, Dai et al., 1995), although DA levels in the medial PFC were still found to be

noticeably increased by L-dopa (Carey, Pinheiro-Carrera, Dai, Tomaz, & Huston, 1995; see also Antinori et al., 2018; Devoto et al., 2016). Moreover, two of the reported PD studies further support their interpretation that L-dopa primarily affected striatal DA by showing that the behavioral drug effects were adequately predicted by simulations in a computational model of the basal ganglia (BG) circuit (Frank et al., 2004; Moustafa et al., 2008). In this BG model, unmedicated PD was simulated by a reduced number of intact DA units, such that both tonic and phasic striatal DA activity were reduced. While the reduced phasic DA activity in response to positive outcomes was associated with impaired Go learning in this model, the reduced tonic DA activity was associated with enhanced NoGo learning, both effects paralleling the behavioral results observed in unmedicated PD patients (Frank et al., 2004; Moustafa et al., 2008). In contrast, medicated PD was simulated by a BG model with partly restored striatal DA activity. In this model, increased phasic DA activity (simulating L-dopa effects) led to enhanced Go learning, whereas increased tonic DA activity (simulating the effects of DA agonists often co-administered with L-dopa) led to impaired NoGo learning, together matching the behavioral effects observed in PD patients on versus off medication. Hence, these simulations support the view that the L-dopa effects on positive reinforcement observed in PD patients are specifically driven by increased phasic striatal DA activity in response to positive feedback. Transferring these findings to the current study, it might be hypothesized that L-dopa strengthened exploitation by enhancing phasic DA activity and prediction error signaling in the striatum, thereby driving choices more strongly towards positive outcomes. It should be kept in mind, however, that the effects of a daily L-dopa treatment in PD patients might not be directly comparable to the effects of a single dose of L-dopa in healthy subjects, given the atypical DA network in the parkinsonian brain and the complex long-term effects of L-dopa (see 1.2.3; Grace, 2008; Hershey, 2003). Still, the assumption of an increased phasic DA activity under L-dopa would also be in line with the above findings of Pessiglione et al. (2006), showing that L-dopa enhances reward prediction error signaling in the striatum of healthy human subjects.

The view that L-dopa primarily affects phasic rather than tonic DA activity is further supported by PET experiments in healthy human subjects (Black et al., 2015; Floel et al., 2008). These studies assessed striatal DA release by use of the radioligand [^{11}C]raclopride (RAC), which specifically binds to D2-like receptors, but is displaced from these receptors when synaptic DA concentration increases (see Egerton et al., 2009; Laruelle, 2000). It was shown that L-dopa (vs. placebo) produces no measurable increase in baseline (tonic) striatal DA release under resting conditions (Black et al., 2015; Floel et al., 2008), but a significant increase in task-evoked (phasic) striatal DA release during motor training (Floel et al., 2008). This increased phasic DA release was associated with improved learning under L-dopa compared to placebo, presumably by enhancing the reinforcing effect of positive feedback during learning (see also Breitenstein et al., 2006; de Vries, Ulte, Zwitserlood, Szymanski, & Knecht, 2010; Frank et al., 2004). Moreover, also *in vivo* microdialysis or voltammetry studies in rats have shown that L-dopa primarily increases phasic (i.e. impulse-dependent) DA release in the striatum (Harun et al., 2016; Keller, Kuhr, Wightman, & Zigmond, 1988; Miller & Abercrombie, 1999; Wightman et al., 1988).

Still, the question remains why L-dopa might specifically enhance phasic and not tonic DA levels, as proposed by these findings. Based on previous research, it has been suggested that exogenous L-dopa is taken up by nigrostriatal DA nerve terminals, converted to DA, stored in synaptic vesicles, and then co-released with endogenous DA upon neural excitation (see 1.2.3; e.g. Breitenstein et al., 2006; Floel et al., 2008; Hälbig & Koller, 2007; Horne et al., 1984). In other words, L-dopa basically "stocks up" presynaptic DA stores in healthy subjects, or "replenishes" these stores in subjects with depleted DA levels like PD patients. From this, it might be assumed that L-dopa boosts DA release in both the phasic and tonic firing mode and hence also increases tonic DA transmission, as often stated in the literature (e.g. Antinori et al., 2018; Guthrie, Myers, & Gluck, 2009; Kroemer et al., 2018; Price, Filoteo, & Maddox, 2009). However, based on the above PET findings, it has been hypothesized that in the intact striatum of healthy subjects, DA catabolism and storage capacity are sufficient to prevent the exogenous DA from being released under resting conditions, i.e. in the tonic state, while it is released during phasic bursts of firing (Tedroff et al., 1996; see also Black et al., 2015; Floel et al., 2008). In line with this notion, studies in healthy human subjects have often reported that L-dopa produces opposite behavioral effects to DA receptor agonists, which are thought to selectively increase tonic (but not phasic) DA signaling (see Breitenstein et al., 2006; Floel et al., 2008; van Eimeren et al., 2009). For example, several studies have shown that L-dopa improves learning in healthy subjects (de Vries et al., 2010; Floel et al., 2005, 2008; Knecht et al., 2004), whereas the DA agonist pergolide was found to impair learning (Breitenstein et al., 2006). Notably, these mixed findings have been attributed to the differential effects of increased phasic (with L-dopa) versus increased tonic (with pergolide) DA signaling, whereby increased tonic activity may actually "mask" phasic DA signals critical for feedback-driven learning (Breitenstein et al., 2006; see also de Vries et al., 2010; van Eimeren et al., 2009). Note that in the parkinsonian brain, however, L-dopa may actually lead to an increase in tonic DA levels due to the reduced storage capacity of DA nerve terminals in the denervated striatum, as suggested by animal studies (see Carey, Dai et al., 1995; Miller & Abercrombie, 1999; Tedroff et al., 1996). Taken together, the reported PET findings and further studies support the view that L-dopa might indeed primarily boost signal-dependent phasic DA release, while leaving tonic DA levels relatively unaffected in healthy subjects.

The assumption that the observed L-dopa effects are primarily attributable to enhanced phasic rather than tonic striatal DA is also in line with studies examining specifically the effects of *tonic* DA modulation on explore/exploit behavior (Beeler et al., 2010; Costa et al., 2014; Humphries et al., 2012). Together, these studies suggest that elevated tonic DA levels in the striatum increase rather than decrease different forms of exploratory behavior, contrary to the L-dopa effect observed here. To better compare and discuss the differential roles of tonic and phasic DA on specific aspects of explore/exploit behavior, it is worth to consider some of the work on tonic DA in more detail. For instance, it was shown that DAT knockdown mice, which are characterized by higher tonic striatal DA levels due to reduced DA reuptake, show increased random exploration as expressed by a higher

softmax β parameter than controls (Beeler et al., 2010). Notably, this result is qualitatively different from the result of the current study, in which random exploration (i.e. the softmax β) was not changed under L-dopa compared to placebo. The view that tonic DA specifically regulates the trade-off between exploitation and *random* exploration is also expressed in the "thrift regulation" hypothesis of DA (see 1.3.1; Beeler, 2012). According to this theoretical framework, tonic DA levels regulate thrift, i.e. the degree to which prior reward learning needs to be exploited to maximize return on energy expenditure. While increased levels of tonic DA reduce thriftiness and facilitate exploratory behavior, decreased levels of tonic DA increase thriftiness and strengthen exploitative behavior. Formally, the framework describes this trade-off also through the softmax β parameter, which reflects how strongly choices are driven by learned reward values. The view that tonic DA regulates action selection via the softmax β parameter is additionally supported by a simulation study on a computational basal ganglia (BG) model (Humphries et al., 2012). This simulation showed that changes in tonic striatal DA levels affect the trade-off between exploitation and random exploration (i.e. the softmax β) on both the neural and behavioral level. More specifically, the study showed that on the neural level, higher tonic DA levels result in a higher β parameter (corresponding to less random exploration) as encoded in the BG output by a more peaked probability distribution for action selection. However, they also showed that on the behavioral level, these associations might paradoxically lead to a situation where high compared to medium tonic DA levels result in more random exploration, consistent with the above findings in hyperdopaminergic DAT knockdown mice (Beeler et al., 2010). Thus, the evidence reported so far suggests that tonic DA in the striatum specifically regulates the trade-off between exploitation and random exploration, which was not found to be affected by L-dopa in the present study, suggesting that tonic DA was not significantly affected by the drug. Yet, it should also be noted that the reported studies on tonic DA all modeled choice behavior with the standard softmax function (i.e. without exploration bonus) and could thereby only capture tonic DA effects on random but not directed exploration. Hence, it may not be inferred from these studies how tonic DA activity affects the explore/exploit trade-off when directed exploration is additionally accounted for.

The effects of tonic DA modulation on *directed* exploration were, however, investigated in an animal study by Costa et al. (2014). In this study, the choice behavior of monkeys was examined in a probabilistic reinforcement learning task after systemic administration of the selective DAT inhibitor GBR-12909, which is known to increase tonic striatal DA levels by slowing DA reuptake (Zhang, Doyon, Clark, Phillips, & Dani, 2009). It was found that DAT blockade *increased* directed (novelty-driven) exploration, i.e. biasing the monkeys to select novel options over familiar ones. Cognitive modeling revealed that increased exploration after DAT blockade was driven by the assignment of a *higher* subjective value to novel (i.e. more uncertain) choice options. In contrast, DAT blockade did not significantly change the trade-off between exploitation and *random* exploration, as expressed by the softmax β parameter. Hence, these behavioral effects after DAT blockade show basically the same pattern – albeit in the opposite direction – as the L-dopa effects observed in the current study. Here,

directed exploration was *reduced* under L-dopa, driven by a *lower* subjective value assigned to more uncertain options (i.e. lower φ), while random exploration (i.e. the softmax β) was also not affected. To explain these opposite results, it might be assumed that the behavioral effects produced by DAT blockade are not solely attributable to increased tonic striatal DA levels, but also to additional changes within the DA system. First, increased tonic striatal DA levels are considered to reduce phasic striatal DA release through feedback inhibition via presynaptic DA autoreceptors (Bilder et al., 2004; Cools, 2006; Floresco et al., 2003; Ford, 2014; Grace, 1991, 2000). Hence, DAT blockade should result in both increased tonic and reduced phasic striatal DA levels. Indeed, DAT knockdown mice were shown to exhibit not only elevated tonic striatal DA levels, but also a clear (ca. 75 %) reduction in the amplitude of phasic striatal DA release (Zhuang et al., 2001). In contrast, L-dopa is known to *increase* the amplitude of phasic striatal DA release (see above; e.g. Harun et al., 2016; Keller et al., 1988; Miller & Abercrombie, 1999; Wightman et al., 1988), which might already (partly) explain the opposite results of both studies. Moreover, it might be assumed that also changes in the prefrontal DA system might have contributed to the observed effects on directed exploration in both studies. In particular, the relative balance between striatal and prefrontal DA activity might play a crucial role in regulating the trade-off between exploitation and *directed* exploration, as discussed in more detail in the following section (see 6.2.1.2). To briefly outline this idea here with respect to the current discussion: A decrease in directed exploration – as observed after L-dopa administration – might result from a state of *high phasic striatal* versus *low prefrontal* DA transmission. In contrast, an increase in directed exploration – as observed after DAT blockade – might result from a state of *low phasic striatal* versus *high prefrontal* DA transmission. Note that prefrontal DA transmission may indeed be increased after DAT blockade, either in relative terms (i.e. relative to the reduced phasic striatal DA activity, see above) or in absolute terms. The latter assumption would be supported by studies in rodents showing that DAT inactivation by knockout or GBR-12909 significantly increases extracellular DA levels in the PFC (Carboni, Silvagni, Vacca, & Di Chiara, 2006; see also Bai et al., 2014; Xu et al., 2009). Since DAT is also substantially expressed in the PFC of monkeys and humans (Ciliax et al., 1999; Lewis et al., 2001), DAT blockade by GBR-12909 might therefore indeed have led to an absolute increase in prefrontal DA levels in the money study by Costa et al. (2014), contributing to a shift in the striatal/prefrontal DA balance. In conclusion, the reported findings from animal and BG network studies suggest that the behavioral effects of tonic DA modulation are different – and partly opposite – from the L-dopa effects observed in the current study: While increased tonic DA levels (probably accompanied by reduced phasic striatal DA release) were shown to promote random and/or directed exploration over exploitation, L-dopa was found to promote exploitation over directed exploration. The discrepancies in these findings, together with the PET studies reported above, support the idea that the observed L-dopa effects primarily resulted from increased phasic striatal DA activity, and that tonic striatal DA levels might not have been substantially altered by L-dopa.

Finally, the assumption that L-dopa strengthened exploitation by increasing phasic striatal DA activity are also in line with studies examining the role of striatal DA on temporal discounting. Temporal discounting, or delay discounting, refers to the phenomenon that subjective reward values are discounted by imposed time delays (see Green & Myerson, 2004). While the concepts of temporal discounting and explore/exploit behavior describe different aspects of reward-based decision making, they might still be strongly related. Remember that exploitation is commonly defined as choosing the option with the highest (immediate) expected reward value. Accordingly, the current study showed that exploitative choices were associated with higher expected reward values than exploratory choices (see Table A4 in the appendix). Thus, it might be assumed that exploitation reflects a more short-sighted (impulsive) choice behavior driven towards immediate rewards, whereas exploration reflects a more far-sighted choice behavior to maximize reward in the long term. Thereby, greater temporal discounting and impulsivity might be linked to a more exploitative choice behavior. Notably, a previous fMRI study (Pine et al., 2010) on temporal discounting showed that a single dose of L-dopa (vs. placebo) increased impulsivity in healthy human subjects, driving choices more strongly towards immediate compared to delayed monetary rewards. Moreover, the study also showed that the greater choice impulsivity under L-dopa was associated with a corresponding increase in the striatal correlates for temporal discounting. The idea that L-dopa enhances choice impulsivity via increased striatal DA activity is further supported by several studies investigating DA medication effects in Parkinsonism. For example, it was shown that PD patients under DA treatment become prone to impulse control disorders (ICDs) like pathological gambling, compulsive shopping, or binge eating (see O'Sullivan et al., 2009; Weintraub, 2008), and that both L-dopa and DA agonists are independently associated with the development of ICDs (Weintraub et al., 2010). Moreover, experimental studies in PD patients showed that DA treatment with L-dopa or DA agonists leads to an increase in choice impulsivity and reward-related striatal activity (Cools, Barker, Sahakian, & Robbins, 2003; Housden, O'Sullivan, Joyce, Lees, & Roiser, 2010; Voon, Pessiglione et al., 2010; Voon, Reynolds et al., 2010). These effects were especially observed in PD patients susceptible to ICDs, suggesting that drug-induced alterations in striatal DA activity may play a central role in the development of impulsive behaviors in PD. In addition, animal studies showed that both chronic L-dopa treatment in parkinsonian rats as well as acute L-dopa treatment in control rats lead to an increase in impulsive behavior (Carvalho et al., 2017), the latter being consistent with the finding that acute L-dopa administration increases choice impulsivity in healthy human subjects (Pine et al., 2010). Taken together, these studies show that L-dopa administration enhances choice impulsivity in both healthy subjects and PD patients, presumably mediated by increased striatal DA release.

The idea that increased striatal DA release promotes impulsive behavior is also supported more directly by human PET studies. For example, it was shown that higher trait impulsivity is associated with increased striatal DA release in healthy subjects, probably resulting from diminished feedback inhibition of DA release due to a lower D2/D3 autoreceptor availability in impulsive subjects (Buckholtz

et al., 2010; see also Dalley et al., 2007; Lee et al., 2009). Similarly, increased striatal DA release was also found in PD patients with DA medication-induced impulsive or compulsive disorders compared to PD patients without such disorders (Evans et al., 2006; Steeves et al., 2009). Although the exact mechanisms by which striatal DA release mediates impulsive behaviors are currently not understood, it might be assumed that the striatal DA system plays a critical role in the differential valuation of immediate and delayed rewards. Animal studies have shown, for example, that behavioral discounting is associated with a decreased striatal DA response for delayed versus immediate rewards during the decision process (Fiorillo, Newsome, & Schultz, 2008; Kobayashi & Schultz, 2008; Schultz, 2010) and have also established the causal relationship between the magnitude of the striatal DA response and the behavioral choice preference in intertemporal decision making (Saddoris et al., 2015). These studies further showed that both longer delays and lower reward magnitudes lead to a similar decrease in the striatal DA response, suggesting that temporal delays influence choices by reducing the apparent subjective value for delayed rewards (see Schultz, 2010). In line with this notion, evidence from human fMRI studies indicates that choosing between immediate and delayed rewards involves the comparison of neurally encoded subjective values within a valuation network including striatal regions (Kable & Glimcher, 2007; Peters & Büchel, 2011). In particular, it was shown that a greater preference for immediate over delayed rewards (i.e. steeper discounting) is associated with an increased reward-related activity in the ventral striatum, suggesting that high impulsivity may be linked to a striatal hypersensitivity for rewards (Hariri et al., 2006). Transferring these findings to the current study, it could be hypothesized that by increasing phasic striatal DA release, L-dopa might have enhanced the overvaluation of immediate compared to delayed rewards, thereby leading to a more impulsive, exploitative choice behavior. However, it should be noted that this assumption only focuses on L-dopa effects on the striatal DA system, while the neural correlates of temporal discounting and its modulation are surely more complex than that, also involving other brain regions such as the vmPFC, OFC, and PCC (see Kable & Glimcher, 2007; Peters & Büchel, 2011). To conclude, the findings on temporal discounting are largely consistent with those on explore/exploit behavior reported above, both supporting the view that L-dopa strengthens the positive reinforcing effect of immediate rewards via increased phasic DA release in the striatum, thereby fostering impulsive and exploitative choice behavior.

6.2.1.2 L-dopa effects with respect to prefrontal DA function

While the previous section mainly focused on potential L-dopa effects on striatal DA activity, striatal and prefrontal DA systems are known to strongly interact and to be both involved in explore/exploit behavior. While the striatal DA system has been implicated in the trade-off between exploitation and random exploration, the *prefrontal* DA system has been proposed to promote directed (uncertainty-driven) exploration, as reviewed in the introduction (see 1.3). To briefly summarize these findings, genetic and pharmacological studies have shown that prefrontal DA function positively predicts uncertainty-driven exploration (Blanco et al., 2015; Frank et al., 2009; Kayser et al., 2015) and related

functions such as behavioral flexibility (e.g. Egan et al., 2001; Fallon et al., 2015; Malhotra et al., 2002) and risk-seeking behavior (Lancaster et al., 2012). Also, uncertainty-driven exploration was found to be reduced in schizophrenic patients and to negatively correlate with the severity of anhedonia, a negative symptom that has been linked to degraded prefrontal DA function (Strauss et al., 2011). Furthermore, human brain stimulation experiments point towards a causal role of the prefrontal cortex, especially frontopolar regions, in promoting uncertainty-driven exploration (Raja Beharelle et al., 2015; Zajkowski et al., 2017), although the DA specificity of these effects remains to be shown.

In the current study, directed exploration – as expressed by the φ parameter – has been shown to be reduced under L-dopa compared to placebo. Since previous research suggests that higher prefrontal DA function is associated with increased directed exploration, this finding might be interpreted in the way that L-dopa reduced directed exploration by reducing prefrontal DA activity. However, this interpretation seems relatively unlikely at first sight, since L-dopa, as a metabolic precursor of DA, might be assumed to elevate DA levels in both striatal and prefrontal regions, as it was previously shown in rats (Carey, Dai et al., 1995; Carey, Pinheiro-Carrera et al., 1995). Yet, these studies also showed that the DA increase induced by L-dopa is considerably (i.e. 50-60 times) larger in striatal than in prefrontal brain regions. Thus, an alternative interpretation of the findings might be that L-dopa reduced prefrontal DA activity *relative* to striatal DA, thereby leading to a corresponding shift in the explore/exploit trade-off towards more exploitation and less directed exploration. The assumption that L-dopa primarily increases striatal over prefrontal DA function is further supported by previous research suggesting that the drug mainly acts on the striatal level, as discussed above (see 6.2.1.1).

To better understand how a relative shift in the striatal/prefrontal DA balance might affect explore/exploit behavior, it should be considered that frontostriatal interactions are believed to play a crucial role in regulating this trade-off. Specifically, it is assumed that uncertainty-driven exploration is implemented on the neural level via a frontostriatal top-down control mechanism, as already described in the introduction (see 1.1.4). According to this idea, the FPC tracks the relative uncertainty of alternative choice options and may interfere via frontostriatal connections with the striatal DA system to override exploitative choice tendencies and trigger exploration. Empirical support for this idea is provided by human brain stimulation studies, showing that the FPC plays a causal role in promoting exploratory behavior (Raja Beharelle et al., 2015), especially uncertainty-driven exploration (Zajkowski et al., 2017). Furthermore, the idea that frontostriatal interactions play a crucial role in the explore/exploit trade-off is also supported by a recent fMRI functional connectivity study in humans (Morris et al., 2016). The study showed that resting-state functional connectivity between the FPC and ventral striatum correlates positively with uncertainty-driven exploration as assessed with the clock task. Additionally, animal experiments provide more direct evidence that reward-related behavior is regulated by a DA-dependent frontostriatal top-down control mechanism. For instance, a PET study in monkeys showed that prefrontal DA depletion leads to an increase in striatal DA release and greater reinforcement sensitivity, i.e. an improved ability to learn from rewarding feedback (Clarke et al.,

2014). Moreover, reinforcement sensitivity was shown to correlate with the PET measure of striatal but not prefrontal DA activity, suggesting that the behavioral effect was specifically mediated by increased striatal DA release. Similar evidence is provided by a rat study showing that PFC lesions induce greater impulsivity, which is in turn alleviated by striatal administration of the D2 receptor antagonist sulpiride, in line with a frontostriatal top-down control of impulsive behavior (Pezze, Dalley, & Robbins, 2009). Finally, the notion of a balanced interaction between the striatal and prefrontal DA system also corresponds well to the "neurochemical reciprocity between DA in the PFC and the striatum" (Cools & D'Esposito, 2011, p. e117), according to which prefrontal and striatal DA function are inversely related (e.g. Akil et al., 2003; Meyer-Lindenberg et al., 2005; Pycock, Kerwin, & Carter, 1980; Roberts et al., 1994; see also review by Cools & D'Esposito, 2011).

On the basis of these findings, it might be hypothesized (in a simplified way) that exploration and exploitation are associated with different functional brain states, characterized by a reciprocal relationship between striatal and prefrontal DA activity: On the one hand, a state of *high phasic striatal* versus *low prefrontal* DA activity, which is associated with enhanced value-driven, exploitative behavior. On the other hand, a state of *low phasic striatal* versus *high prefrontal* DA activity, which is associated with enhanced uncertainty-driven exploration. Further, it might be assumed that the shift between both states is adaptively regulated via frontostriatal interactions, e.g. through increased or decreased prefrontal top-down control. Based on this idea, a change in the relative balance between striatal and prefrontal DA activity should shift the explore/exploit trade-off accordingly, leading to a higher tendency to exploit or to explore. More specifically, a relative increase in prefrontal over phasic striatal DA activity (as assumed under DAT blockade, see 6.2.1.1) should favor the state in which the prefrontal system dominates, leading to more directed exploration and less exploitation – as observed by Costa et al. (2014). On the other hand, a relative increase in phasic striatal over prefrontal DA activity (as assumed under L-dopa) should strengthen the state in which the striatal system dominates, stabilizing it against prefrontal control and leading to more exploitation and less directed exploration – as observed in the current study.

Consistent with this assumption, several findings on explore/exploit behavior may be interpreted in terms of a relative shift in the striatal/prefrontal DA balance. For instance, it was found that uncertainty-driven exploration is reduced in COMT Val/Val compared to Met/Met subjects (Frank et al., 2009), whereby the Val allele has been linked to both lower prefrontal DA levels (Bilder et al., 2004) and increased phasic (reward-related) striatal DA activity (Brody et al., 2006; Krugel, Biele, Mohr, Li, & Heekeren, 2009). Moreover, uncertainty-driven exploration was found to be reduced in schizophrenic patients compared to healthy controls (Strauss et al., 2011), whereby schizophrenia has also been linked to reduced prefrontal and increased striatal DA function (see 1.2.3; e.g. Abi-Dargham, 2004; Davis et al., 1991; Howes & Kapur, 2009; Lau et al., 2013; Weinstein et al., 2017). Notably, as mentioned above, uncertainty-driven exploration in these patients correlated negatively with anhedonia, a negative symptom attributed to degraded prefrontal DA function (see Abi-Dargham & Moore, 2003;

Brisch et al., 2014; Davis et al., 1991; Strauss et al., 2011), in line with the idea that the behavioral shift in the explore/exploit trade-off may directly scale with the (assumed) shift in the striatal/prefrontal DA balance. Also in line with the view of a reciprocal balance between both DA systems – and of particular interest – are the findings of a human fMRI study on risky decision making, which used a gene composite score across five different DA genes as a positive marker of striatal DA activity (Kohno et al., 2016). The study showed that this positive marker of striatal DA activity was negatively correlated with prefrontal activity during risky decision making. Note that a similar reciprocal relationship has already been found in an earlier human PET study by Kohno et al. (2015), who showed that striatal D2-like receptor availability was positively correlated with reward-related striatal activity, but negatively with risk-related prefrontal activity and risky choice behavior. Interestingly, the authors interpreted these findings as support for the notion that "striatal dopamine signaling modulates top-down corticostriatal control to guide adaptive decision making" (Kohno et al., 2016, p. 701). In other words, these results indicate that striatal DA activity may directly influence prefrontal DA function and reduce its top-down control, according to a frontostriatal bottom-up mechanism. Therefore, individuals with higher striatal DA function may be more reward sensitive and have a less effective cortical inhibition of reward-driven responses, which may lead to a preference for immediate smaller gains over risky delayed ones (see Kohno et al., 2015). With respect to explore/exploit behavior, such a frontostriatal bottom-up mechanism might function to stabilize value-driven exploitation against the initiation of uncertainty-driven exploration (i.e. risky choices). The notion that striatal DA activity can modulate prefrontal DA function is also supported by further research, and different neural mechanisms have been proposed to explain these interactions (see Cools, 2011; Duvarci et al., 2018; Kohno et al., 2015; Simpson & Kellendonk, 2017; Simpson, Kellendonk, & Kandel, 2010). For example, it was shown that increasing striatal DA activity by transient D2 receptor overexpression modulates various aspects of prefrontal DA signaling (Kellendonk et al., 2006; Simpson & Kellendonk, 2017), and that such striatal-to-prefrontal interactions might be mediated by the degree of long-range neural synchrony between VTA DA neurons and the PFC (Duvarci et al., 2018). Aside from neural synchrony, such interactions might also depend on other regulatory mechanisms within frontostriatal circuits involving, for example, glutamatergic and GABAergic signaling (as discussed by Kohno et al., 2015; see also Seamans & Yang, 2004; Sesack & Grace, 2010). Based on these findings and ideas, it might be speculated that such frontostriatal bottom-up interactions could also occur under pharmacological stimulation of the striatal DA system by L-dopa, potentially even mediated by drug-induced changes in neural synchrony, which have been observed in response to both L-dopa and haloperidol (see below, 6.2.1.3). Hence, this might provide an explanation how L-dopa, by boosting striatal DA activity, could indeed have reduced prefrontal control and thereby uncertainty-driven exploration without actually reducing prefrontal DA concentrations in absolute terms. However, further research would be needed to support these ideas and to investigate the neural mechanism underlying such a reciprocal regulation of the explore/exploit trade-off.

In conclusion, previous research suggests that striatal and prefrontal DA fulfill different and even opposing functions within the explore/exploit trade-off: While the striatal DA system – especially with its phasic activity – promotes impulsive and exploitative behavior, the prefrontal DA system is thought to promote uncertainty-driven exploration via a frontostriatal top-down mechanism. Since L-dopa is considered to primarily enhance phasic striatal DA activity, it should shift the balance between both systems towards a state more strongly dominated by striatal DA, hence explaining the observed behavioral shift towards more exploitation and less directed exploration. Further research is needed, though, to elucidate the precise neural mechanisms behind these effects and to disentangle the differential contributions of striatal vs. prefrontal and phasic vs. tonic DA activity in the regulation of explore/exploit behavior.

6.2.1.3 L-dopa effects with respect to further aspects of DA function

Finally, it should be noted that the actual neural mechanism behind the observed L-dopa effects might be much more complex than suggested by the studies discussed so far. Previous research has shown that there are various other aspects of DA function, in addition to the striatal/prefrontal and phasic/tonic aspect, which might have contributed to the observed L-dopa effects.

One of these aspects relates to the fact that DA signaling depends on different DA receptor subtypes, i.e. D1-like vs. D2-like receptors, which were shown to exert different functions in driving reward-based and risky decision making (e.g. Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Floresco, 2013; Keeler, Pretsell, & Robbins, 2014; Kravitz, Tye, & Kreitzer, 2012; Simon et al., 2011; St Onge et al., 2011). Reviewing this literature shows, however, that the reported relationships between D1/D2 signaling and behavior are highly complex and often inconsistent between studies, not allowing to conclusively relate these findings to the observed L-dopa effects in the current study. For example, D1 and D2 receptors are differentially expressed in distinct pathways within the BG circuit, which are assumed to play different roles in positive and negative reinforcement: a direct ("Go") pathway, which predominantly expresses excitatory D1 receptors and mediates positive reinforcement, and an indirect ("NoGo") pathway, which mainly expresses inhibitory D2 receptors and mediates negative reinforcement (see Bromberg-Martin et al., 2010; Frank et al., 2004; Frank & O'Reilly, 2006; Kravitz et al., 2012). In line with this notion, evidence from a human PET study (Cox et al., 2015) indicates that phasic striatal D1 receptor signaling correlates with positive reinforcement, whereas striatal D2 receptor signaling relates to negative reinforcement in an inverted-U-shaped fashion. Based on these functional distinctions, it might be speculated that the observed increase in exploitative behavior under L-dopa is primarily attributable to enhanced phasic D1 receptor signaling in the direct pathway implicated in positive reinforcement. However, there is also evidence for an involvement of the D2 receptor in positive reinforcement and exploitative behavior. For example, another human study (Eisenegger et al., 2014) found that while D2 receptor blockade does not affect the learning rate for positive reinforcement, it still affects choice behavior as shown by a reduced softmax β parameter

indicating less value-driven exploitation. Furthermore, also prefrontal D1 and D2 receptors have been linked to explore/exploit behavior, as they were shown to modulate risk/reward decision making, albeit in a complex and often nonlinear fashion (Floresco, 2013; Simon et al., 2011; St Onge et al., 2011). For example, one rat study showed that D1 and D2 receptors make dissociable and partly complementary contributions to risk/reward judgments in decision making, from which it was concluded that the explore/exploit trade-off may critically depend on a "fine balance between D1/D2 receptor activity" (St Onge et al., 2011, p. 8625). Altogether, the reported studies show that distinguishing D1 and D2 receptor function in different DA systems seems crucial to understand how DA regulates the explore/exploit trade-off on the neural level. Yet, the methodology of the current study does not allow to unequivocally infer how L-dopa affected signaling in the different DA receptor subsystems and how these changes might have contributed to the observed behavioral effects.

Another notable aspect of DA signaling, which has been implicated in explore/exploit behavior, is the degree of neural synchrony within dopaminergic networks. Previous research suggests that adjacent DA neurons show electrical coupling, which can lead to synchronous firing within a population of DA neurons (see Grace et al., 2009; Grace & Bunney, 1983). Moreover, there is by now strong evidence that the degree of neural synchrony within the BG network depends on nigrostriatal DA levels (Bergman et al., 1994, 1998; Fountas & Shanahan, 2017; Plenz & Kital, 1999) and is modulated by DA drugs like L-dopa and haloperidol (Brown et al., 2001; Burkhardt, Constantinidis, Anstrom, Roberts, & Woodward, 2007; Ruskin et al., 1999; Yael et al., 2013; see also reviews by Brittain & Brown, 2014; Hammond, Bergman, & Brown, 2007; Quiroga-Varela et al., 2013). Crucially, it has also been suggested that the degree of neural synchrony within different DA networks might play an important role in the regulation of explore/exploit behavior. For instance, a simulation experiment on a spiking BG network model found that striatal DA levels regulate explore/exploit behavior by changing the level of neural synchrony within the BG circuit (see 1.3.1; Mandali et al., 2015). More specifically, this study showed that high (vs. intermediate) striatal DA levels were associated with a more exploitative choice behavior in a restless four-armed bandit task, which was mediated by decreased neural synchrony. Assuming that the L-dopa condition (vs. placebo) in the current study corresponds to the high (vs. intermediate) striatal DA condition in the simulation experiment, the predictions from the BG model are broadly in line with the finding that L-dopa increased exploitation in the bandit task. Note, however, that both findings are not directly comparable, as the simulation study focused specifically on the subcortical substrates of explore/exploit behavior and used a choice model that only captures random but not directed exploration. Aside from this study, striatal DA was furthermore shown to modulate the degree of neural synchrony not only within the BG circuit, but also in DA networks extending to prefrontal cortical regions. For example, a recent transgenic mice study (Duvarci et al., 2018) showed that striatal DA hyperfunction leads to alterations in the long-range neural synchrony between VTA DA neurons and the PFC, which are associated with an impairment in PFC-dependent cognitive functions. As discussed above, this finding points to a potential mechanism by which striatal DA could influence

prefrontal DA function in order to control the explore/exploit trade-off from "bottom-up", which might have contributed to the observed shift in explore/exploit behavior under L-dopa. Yet, within the confines of the current study design, it cannot be answered if and how drug-induced changes in neural synchrony might have contributed to the observed behavioral effects.

In conclusion, various aspects of DA signaling may have been involved in producing the observed L-dopa effect in the current study, and further research is needed to examine their potential roles in the explore/exploit trade-off. Based on the current knowledge, however, most evidence supports the (simplified) conclusion stated above, which attributes the observed L-dopa effect mainly to increased phasic striatal DA activity and a relative shift in the striatal/prefrontal DA balance, resulting altogether in a strengthening of value-driven exploitation against uncertainty-driven exploration.

6.2.2 Haloperidol effects

In contrast to L-dopa, haloperidol was not found to substantially affect explore/exploit behavior across subjects. As shown by cognitive modeling, the drug did not shift the group-level φ or β parameter relative to placebo. On the subject level, however, haloperidol was found to elicit a complex pattern of mixed drug effects on both the φ and β parameter. Additionally, haloperidol showed a tendency to reduce the group-level variance for the φ parameter (Λ^{φ}) relative to placebo and L-dopa. The following discussion of these results will first focus on haloperidol effects on the striatal level, before potential drug effects on the prefrontal level will be considered.

6.2.2.1 Haloperidol effects with respect to striatal DA function

The absence of a clear behavioral effect under haloperidol is partly surprising, given previous findings on the dopaminergic modulation of explore/exploit behavior as discussed in the preceding section (see 6.2.1). As a potent D2 receptor antagonist, haloperidol would be expected to reduce DA transmission and thereby modulate DA-dependent behaviors, including the trade-off between exploitation and exploration. According to the initial hypothesis, a haloperidol-induced reduction of tonic striatal or prefrontal DA function would be expected to decrease exploratory behavior relative to placebo (see 1.4.2). Yet, the findings discussed above (see 6.2.1.1) also allow for an alternative hypothesis, according to which haloperidol might exhibit an effect opposite to L-dopa and reduce reward-related (phasic) striatal activity, thereby reducing exploitation and facilitating exploration. Consistent with the latter assumption, previous experimental studies have shown that haloperidol and other D2 receptor antagonists reduce reward-driven exploitative behavior in healthy human subjects (Eisenegger et al., 2014; Pessiglione et al., 2006; Pleger et al., 2009). For example, it was found that haloperidol (1 mg), compared to L-dopa (100 mg), reduces exploitative choice behavior and striatal reward prediction error signals in a reinforcement learning task with monetary gains and losses (Pessiglione et al., 2006). This drug effect was specifically observed for positive reinforcement ("gain condition"), but not for negative reinforcement ("loss condition"). To provide further insights into the nature of this effect, the

study combined the fMRI results with a cognitive modeling approach, as described in more detail above (see 6.2.1.1). They inferred from this analysis that haloperidol reduced the apparent value of a monetary reward from £1.0 (i.e. the placebo reference value) to a value of £0.71, which could in turn explain the reduced exploitative choice behavior under haloperidol observed in their study. Hence, these results suggest that haloperidol, by decreasing striatal DA activity, reduces the apparent value of rewards and thus the tendency to exploit these rewards. Consistent with these findings, another human fMRI study showed that both reward-driven choice behavior and reward-related striatal activity are enhanced under L-dopa and reduced under haloperidol in a tactile decision making task (Pleger et al., 2009). Likewise, also the D2 receptor antagonist sulpiride was shown to reduce exploitative choice behavior and reward-related striatal activity in healthy human subjects (Eisenegger et al., 2014; McCabe, Huber, Harmer, & Cowen, 2011). For example, Eisenegger et al. (2014) found that sulpiride impairs exploitative choice behavior in the gain domain, but not in the loss domain, consistent with the above haloperidol findings in the same behavioral paradigm (Pessiglione et al., 2006). In sum, these findings support the notion that haloperidol – opposite to L-dopa – decreases reward-related striatal DA activity and positive reinforcement, thereby shifting the trade-off between exploitation and exploration towards the latter. Yet, no such behavioral (or neural) effects have been observed under haloperidol treatment in the current study (see also fMRI discussion in section 6.3.2), raising the question how the absence of such effects might be explained.

Importantly, however, numerous studies yielded results contrary to the findings reported above, showing that a single dose of haloperidol (or other D2 antagonists) does not always produce the expected antidopaminergic effects. In fact, several studies found either no clear haloperidol effect on the behavioral or neural level, or even reported that haloperidol paradoxically enhances rather than impairs DA signaling and DA-dependent behaviors. For the following discussion, it is worth to consider some of these studies in more detail.

To begin with, a large number of animal studies have found that acute low doses of haloperidol and other D2 antagonists actually *stimulate* DA signaling, contrary to the antidopaminergic effects observed under chronic and high-dose treatment with these drugs (as reviewed by Frank & O'Reilly, 2006; Knutson & Gibbs, 2007; Starke et al., 1989). It is generally assumed that these opposite drug effects result from the fact that D2 agents can act on both post- and presynaptic D2 receptors: While blockage of postsynaptic D2 receptors reduces DA signaling and exerts antidopaminergic effects, the blockage of presynaptic D2 autoreceptors is thought to stimulate (phasic) DA signaling due to reduced feedback inhibition of DA release (see Ford, 2014; Frank & O'Reilly, 2006; Schmitz et al., 2003). Crucially, it has been shown that low doses of D2 agents primarily exhibit presynaptic effects *in vivo*, potentially reflecting the higher binding affinity of pre- versus postsynaptic D2 receptors or the larger D2 receptor reserve at presynaptic sites (see Ford, 2014; Knutson & Gibbs, 2007; Meller, Bohmaker, Namba, Friedhoff, & Goldstein, 1987; Neve & Neve, 1997). Indeed, evidence from numerous *in vivo* microdialysis and voltammetry studies in rodents and nonhuman primates support this notion,

showing that single low doses of haloperidol and other D2 antagonists potentiate phasic DA release in the striatum (e.g. Garris et al., 2003; Jaworski, Gonzales, & Randall, 2001; Kuroki, Meltzer, & Ichikawa, 1999; Moghaddam & Bunney, 1990; Pehek, 1999; Schwerdt et al., 2017; Westerink et al., 2001; Westerink, 2002; Wu et al., 2002; Youngren, 1999), whereas single low doses of D2 agonists diminish it (e.g. Kennedy, Jones, & Wightman, 1992; Stamford, Kruk, & Millar, 1991; see also review by Starke et al., 1989). Similarly, pharmacological MRI studies in animals have shown that amphetamine-induced increases in striatal blood volume are potentiated by low doses of D2 antagonists (Chen, Choi, Andersen, Rosen, & Jenkins, 2005; Schwarz et al., 2004) and blunted by low doses of D2 agonists (Chen et al., 2005), consistent with the notion of a presynaptic drug mechanism. Also in humans, acute low doses of haloperidol were found to increase resting cerebral blood flow in the striatum of healthy subjects, in line with a DA-stimulating drug effect (Handley et al., 2013). Together, these findings suggest that due to presynaptic drug actions, D2 agents can affect DA signaling in the exact opposite way than commonly assumed when taking only postsynaptic mechanisms into account (as in the initial hypothesis, see above).

In line with these findings, several human studies have reported “paradoxical” drug effects from single low doses of haloperidol or other D2 antagonists, which are clearly in accordance with the notion of a presynaptic DA-stimulating drug mechanism. For example, it was shown that an acute dose of 2 mg haloperidol, i.e. the same dose as used in the current study, enhances positive reinforcement (Go learning), but impairs negative reinforcement (NoGo learning) relative to placebo in healthy subjects (Frank & O'Reilly, 2006). Interestingly, this haloperidol effect showed exactly the same pattern as the behavioral effect seen in medicated PD patients compared to healthy controls with the same task (Frank et al., 2004), suggesting that low doses of haloperidol can exert similar DA-stimulating effects as the DA medication used for PD therapy (i.e. L-dopa and DA agonists). Furthermore, these paradoxical haloperidol effects were specifically observed in a subgroup of subjects showing an increase in prolactin levels under haloperidol, which provides an indirect measure of the degree to which the drug increased DA levels via presynaptic mechanisms (see Frank & O'Reilly, 2006; further discussed below with regard to baseline-dependent drug effects). Conclusively, the authors attribute these paradoxical haloperidol effects to a presynaptically mediated increase in phasic DA signaling in the BG, which facilitates Go relative to NoGo learning. In addition, they claim that the low dose of 2 mg haloperidol is unlikely to have produced significant postsynaptic effects, which should only be observed at higher drug doses and/or chronic administration. They support this claim by arguing that a substantial blockade of postsynaptic D2 receptors would be expected to produce sedative and Parkinson-like side effects, leading to a slowing of reaction times (see Sanberg, 1980). Yet, a slowing of reaction times has not been observed in their study or other studies using similar low doses of D2 antagonists (e.g. Mehta, Manes, Magnolfi, Sahakian, & Robbins, 2004; Mehta, Sahakian, McKenna, & Robbins, 1999; Peretti et al., 1997), including the current one. Consistent with these haloperidol findings, it was shown that low doses of the D2 antagonist sulpiride enhance reward versus

punishment learning in healthy human subjects, which was associated with a drug-induced increase in striatal prediction error signaling (van der Schaaf et al., 2014). Likewise, another study found that the same low dose of sulpiride enhances reinforcement learning in healthy subjects relative to placebo (Mehta, Hinton, Montgomery, Bantick, & Grasby, 2005). Furthermore, also low doses of amisulpride, another D2 antagonist, have been shown to enhance reward-based decision making and striatal prediction error signaling in healthy volunteers (Jocham, Klein, & Ullsperger, 2011). In contrast, low doses of the D2 agonist pramipexole were found to impair rather than enhance reward responsiveness and reward-based decision making in healthy participants (Pizzagalli et al., 2008). In sum, all these behavioral and neural drug findings support the notion that D2 agents, when acutely administered at low doses, primarily act via presynaptic mechanisms, thereby leading to seemingly paradoxical effects in which DA signaling is stimulated by DA antagonists and reduced by D2 agonists.

In addition to these paradoxical drug effects, many studies also found that low doses of haloperidol (i.e. 1-2 mg) either produced no significant drug effect across subjects, or that a clear drug effect was only present in a subgroup of subjects, consistent with the findings of the current study. For example, a null effect of haloperidol was reported by Pine et al. (2010), who conducted a human fMRI experiment with a similar placebo-controlled, counterbalanced, within-subject design as employed in the current study. In that study, they did not find a significant difference between haloperidol (1.5 mg) and placebo on impulsive behavior or its neural correlates, whereas L-dopa (150 mg) vs. placebo showed clear effects on both the behavioral and neural level (see also 6.2.1.1). In addition, also the study of Pessiglione et al. (2006), which examined the effects of haloperidol (1 mg) and L-dopa (100 mg) on reinforcement learning in a placebo-controlled, between-subject design, did not find a significant behavioral effect of haloperidol relative to placebo. In their study, only the statistical tests of haloperidol against L-dopa yielded significant effects on choice behavior. Moreover, comparing both of their drug conditions separately against placebo reveals that L-dopa exhibited a much stronger (and statistically significant) effect on choice behavior than haloperidol, which suggests that the reported effect between both drugs was primarily driven by a strong L-dopa effect, as observed in the current study. In addition to these null findings, other studies showed that the effects of low doses of haloperidol and other D2 agents strongly depend on individual baseline measures (see below and 1.2.4). Crucially, this baseline-dependency might also explain why some studies found no overall drug effect across subjects or only observed drug effects in a specific subgroup of subjects. For instance, the aforementioned study by Frank and O'Reilly (2006) showed that the extent to which haloperidol affects reinforcement learning strongly depends on the individual working memory span, which is mostly taken as an indicator of baseline DA levels (but see below). After median splitting their sample, they found a significant haloperidol effect only for subjects with low working memory span, but no effect for subjects with high working memory span. Similar span-dependent drug effects were also found for other D2 antagonists and agonists on both the behavioral and neural level (e.g. Gibbs & D'Esposito, 2005; Kimberg et al., 1997; van der Schaaf et al., 2014; see also review by Cools & D'Esposito, 2011).

Interestingly, it has been proposed by Frank and O'Reilly (2006) that these span-dependent differences might not necessarily reflect individual differences in baseline DA levels, as commonly assumed (see 1.2.5; e.g. Cools & D'Esposito, 2011), but rather individual differences in the sensitivity to D2 receptor stimulation. In line with this idea, pharmacogenetic studies have shown that the extent to which D2 drugs affect behavior depends on genetic differences in the D2 receptor system (e.g. Cohen, Krohn-Grimberghe, Elger, & Weber, 2007; Eisenegger et al., 2014; Kirsch et al., 2006). These studies showed, for instance, that D2 drugs affect reward-related brain activity and behavior specifically in a subgroup of human subjects carrying the A1 allele of the Taq1A polymorphism, which is associated with a substantial reduction in the availability of striatal D2 receptors (Gluskin & Mickey, 2016; Jönsson et al., 1999; Pohjalainen et al., 1998; Thompson et al., 1997). Hence, these findings demonstrate that the sensitivity to D2 drug effects strongly depends on individual differences in the D2 receptor system, for which the working memory span might be (partly) predictive. Furthermore, it was also proposed that such individual differences in the D2 receptor system might strongly influence the extent to which a specific dose of a D2 drug acts on pre- versus postsynaptic receptors (see Cools et al., 2009; Frank & O'Reilly, 2006; van der Schaaf et al., 2014). For instance, Frank and O'Reilly (2006) hypothesized that low-span subjects may have a higher sensitivity to D2 receptor stimulation and are therefore more susceptible to presynaptic drug effects from low doses of D2 agents than high-span subjects, consistent with the behavioral results of their experiment. Further research is needed, though, to actually reveal how D2 receptor sensitivity and other factors determine the extent to which a specific drug dose acts on pre- versus postsynaptic receptors and how these cellular actions relate to individual differences in the magnitude and direction of behavioral drug effects. Taken together, the reported findings indicate that various factors influence the specific pattern of D2 drug effects, including the exact drug, the dosage, the examined behavior, and individual baseline measures, which might explain the high variability of drug effects within and across studies and also the occasional null findings observed in some studies, including the current one.

Based on these findings, it may be concluded that the behavioral null effect of haloperidol in the current study can be explained from a number of factors. First, it might be assumed that the administered dose of 2 mg haloperidol was too low to substantially block postsynaptic D2 receptors and produce the expected antidopaminergic effects, e.g. a reduction in phasic striatal DA signaling and exploitation across subjects (according to the later hypothesis, see above). Note that the administered dose of 2 mg is in the range of clinically-prescribed introductory doses and was kept that low in order to minimize the risk of any side effects. However, it has been argued in the literature that D2 agents at such a low dose should primarily exert presynaptic effects, as largely supported by evidence from animal and human studies (see above; e.g. Frank & O'Reilly, 2006; Jocham et al., 2011; Mehta et al., 2005; Pizzagalli et al., 2008; van der Schaaf et al., 2014; although see Eisenegger et al., 2014; Pessiglione et al., 2006; Pleger et al., 2009). Also consistent with this assumption is the observed null-effect of haloperidol on the sEBR, which was used as a positive maker of striatal DA function and should

have been reduced under antidopaminergic drug conditions, as seen for example in schizophrenic patients medicated with haloperidol or other D2 antagonists (see 1.2.5; e.g. Adamson, 1995; Bartkó et al., 1990; Karson, Bigelow, Kleinman, Weinberger, & Wyatt, 1982; see also Jongkees & Colzato, 2016; Kaminer et al., 2011). Second, it might be assumed from previous research that this low dose of haloperidol also exerted presynaptic DA-stimulating effects to some extent, which might have counteracted any postsynaptic drug effects across subjects, leading to the overall behavioral null finding. In particular, it might be assumed that this overall null effect actually reflects a mixture of pre- and postsynaptic effects on the individual level. This assumption is supported by the observation that while no clear drug effect was found for haloperidol on the group level, drug effects of different magnitude and direction were indeed found on the subject level for most of the observed variables, including the exploration bonus parameter φ and the sEBR. Third, it might be assumed that the extent to which haloperidol acted on pre- versus postsynaptic receptors in different subjects may be (in part) determined by the individual D2 receptor system at baseline, as proposed by Frank and O'Reilly (2006; see above). Interestingly, the observation that haloperidol (but not L-dopa) reduced the variance of explore/exploit behavior across subjects (i.e. the Λ^{φ} parameter) might potentially reflect such a baseline-dependency of drug effects. Notably, the pattern of drug effects on the subject level showed a general tendency for haloperidol to increase the φ parameter in low-φ subjects (i.e. low φ at placebo) and decrease it in high-φ subjects. From this pattern, it could be speculated that individual differences in the placebo φ parameter might indicate, to some extent, differential baseline levels of striatal DA function (analogous to the working memory span in the study of Frank and O'Reilly, 2006), which might in turn influence a subject's sensitivity for pre- versus postsynaptic D2 drug effects. For example, it might be speculated that high-φ subjects, who exploit less and might accordingly have lower striatal DA levels at baseline, are specifically susceptible to the DA-stimulating effects of haloperidol, leading to more exploitation (φ reduction) in these subjects. In contrast, low-φ subjects with already high striatal DA levels at baseline might rather show DA-antagonizing effects under haloperidol, leading to less exploitation (φ increase) in these subjects. Still, no such baseline-dependency was observed when taking the working memory span as a predictor of baseline striatal DA function (see 6.4), in contrast to several studies showing span-dependent D2 drug effects (see above; e.g. Frank & O'Reilly, 2006; Gibbs & D'Esposito, 2005; Kimberg et al., 1997; van der Schaaf et al., 2014). Hence, it remains unknown if the observed pattern of haloperidol effects on the subject level indeed reflects some sort of baseline-dependency and, if so, what the neural underpinnings of this effect might be.

Finally, it should be noted that the points discussed so far still leave out several aspects of DA function that might have contributed to the observed pattern of haloperidol effects. These factors include, but are not limited to, prefrontal DA function (see below), the level of phasic vs. tonic DA activity, and the ratio of D1 vs. D2 receptor signaling. For instance, it is possible that aside from the presynaptic increase in phasic DA release, haloperidol might block postsynaptic D2 receptors to some extent even at low doses, thereby leading to a net increase in (phasic) postsynaptic D1 receptor signaling (as discussed by

Kahnt & Tobler, 2017; Shi, Smith, Pun, Millet, & Bunney, 1997; van der Schaaf et al., 2014). Crucially, such a shift in the balance between D1/D2 and phasic/tonic signaling may critically influence the behavioral trade-off between exploration and exploitation (e.g. Beeler, 2012; Burke et al., 2018; Frank, 2005; St Onge et al., 2011). However, it is currently not known how exactly these aspects influence explore/exploit behavior and how they are modulated by D2 agents like haloperidol on an individual level. Future studies will be necessary to resolve these issues, for example by employing a combination of methods such as DA pharmacology, genotyping, and neuroimaging (see e.g. Cohen et al., 2007; Kirsch et al., 2006).

6.2.2.2 Haloperidol effects with respect to prefrontal DA function

Since the discussion above mainly focused on drug effects on the striatal level, the question remains open if and how haloperidol affects prefrontal DA function and how such prefrontal drug actions might have contributed to the observed pattern of behavioral drug effects in the current study. However, the existing literature provides mixed answers to these questions.

On the one hand, a number of findings support the view that single low doses of haloperidol predominantly affect striatal DA, while leaving prefrontal DA function relatively unaffected, as previously discussed by Frank and O'Reilly (2006). First, the expression of D2 receptors in the human brain is several times lower in the PFC than in the striatum (Camps, Cortés, Gueye, Probst, & Palacios, 1989; Hall et al., 1994; Hurd et al., 2001), and it is reasonable to assume that higher doses of D2 agents are required to functionally affect prefrontal activity states (see Frank & O'Reilly, 2006; Seamans & Yang, 2004; Trantham-Davidson, Neely, Lavin, & Seamans, 2004). In particular, D2 autoreceptors were found to be relatively rare in the PFC and other mesocortical projection areas, with some autoreceptor subtypes being completely absent at these sites (Bannon, Wolf, & Roth, 1983; Ford, 2014; Lammel et al., 2008; Roth, 1984; Wolf & Roth, 1990), suggesting that presynaptic effects from low doses of haloperidol are less likely to occur in prefrontal regions. Indeed, several animal experiments have shown that acute low doses of haloperidol selectively stimulate DA release in the striatum, while leaving prefrontal DA levels relatively unaffected (e.g. Kuroki et al., 1999; Pehek, 1999; Rollema, Lu, Schmidt, Sprouse, & Zorn, 2000; Volonté, Monferini, Cerutti, Fodritto, & Borsini, 1997; Westerink, 2002). Moreover, human PET and fMRI studies have provided evidence that D2 antagonists modulate task-related cerebral blood flow and functional connectivity specifically in striatal but not in prefrontal regions (Honey et al., 2003; Mehta et al., 2003). Additionally, D2 drugs were shown to affect working memory processes only when administered systemically, but not when applied locally to the PFC (Arnsten, Cai, Steere, & Goldman-Rakic, 1995; Luciana, 1998; Sawaguchi, 2001; see also Yang & Seamans, 1996). In contrast, prefrontal administration of D1 drugs produces clear effects on working memory processes (Durstewitz & Seamans, 2002; Sawaguchi, 2001; Sawaguchi & Goldman-Rakic, 1991; Williams & Goldman-Rakic, 1995), consistent with evidence that D1 receptors are much more

prevalent in the PFC than D2 receptors (Hall et al., 1994; Hurd et al., 2001; Lidow, Goldman-Rakic, Gallager, & Rakic, 1991).

On the other hand, a number of studies have shown that haloperidol and other D2 drugs can significantly influence prefrontal function. First, *in vivo* microdialysis studies in rats have found that acute systemic administration of haloperidol can stimulate DA release also in the PFC (Hernandez & Hoebel, 1989; Moghaddam & Bunney, 1990; Westerink et al., 2001), contrary to the findings reported above. However, these prefrontal drug effects were shown to strongly depend on the administered dose in a complex fashion. For example, a single dose of 0.5 mg/kg haloperidol was shown to significantly increase prefrontal DA levels (Hernandez & Hoebel, 1989; Moghaddam & Bunney, 1990), whereas no such effect was found with a higher dose of 1.0 mg/kg (Pehek, 1999) or a lower dose of 0.1 mg/kg (Moghaddam & Bunney, 1990). Yet, the higher dose of 1.0 mg/kg was still shown to increase amphetamine-induced DA release in both striatal and prefrontal regions (Pehek, 1999), suggesting that haloperidol at this dose may selectively enhance phasic (stimulus-dependent) but not tonic DA release in the PFC. Aside from these microdialysis results, behavioral studies in rats showed that D2 drugs can produce clear effects on different cognitive functions when selectively applied into the PFC (e.g. Druzin, Kurzina, Malinina, & Kozlov, 2000; Floresco, Magyar, Ghods-Sharifi, Vexelman, & Tse, 2006; St Onge et al., 2011; Zeeb, Floresco, & Winstanley, 2010; see also reviews by Floresco, 2013; Floresco & Magyar, 2006), contrary to the working memory findings reported above (e.g. Sawaguchi, 2001). Yet again, these prefrontal D2 drug effects were found to depend in a complex manner on the exact drug dosage and examined behavior (see Floresco, 2013). For example, while the D2 antagonist eticlopride, injected into the PFC, was shown to affect risky choice behavior and set shifting only at high doses (1.0 µg; Floresco et al., 2006; St Onge et al., 2011), the same drug was found to affect impulsive choice behavior most strongly at medium doses (0.3 µg; Zeeb et al., 2010). Similarly, other D2 drugs applied to the PFC were found to modulate different aspects of working memory performance in a dose-dependent manner, with some aspects already affected at low doses and others only at higher doses (Druzin et al., 2000). Finally, a human study showed that a single oral dose of 3 mg haloperidol administered to healthy subjects not only increases resting-state blood flow in the striatum, but also decreases it in prefrontal regions relative to placebo (Handley et al., 2013; see also Bartlett et al., 1994). It has been proposed, though, that this change in prefrontal blood flow might actually represent a secondary effect that results from primary drug actions on the subcortical level (see Handley et al., 2013). In particular, it was suggested that low doses of haloperidol might stimulate DA signaling primarily in regions with high D2 receptor density, such as the striatum, leading to downstream modulatory processes in brain areas innervated by these D2-dense regions, such as the PFC. In line with this assumption, it was shown that direct injection of the D2 antagonist eticlopride into the VTA of rats affects DA levels not only locally, but also in striatal and prefrontal brain regions in a complex dose-dependent fashion, presumably reflecting downstream regulatory effects within ascending mesolimbic and mesocortical DA pathways (Chen & Pan, 2000). In addition, a single dose of haloperidol (1 mg/kg) was shown to

reduce resting-state functional connectivity between midbrain and prefrontal regions in the rat brain, suggesting that prefrontal haloperidol effects may indeed arise from drug-induced network changes within ascending DA pathways (Gass et al., 2013). Taken together, the reported findings show that D2 drug effects in the PFC, as in the striatum, are highly variable and critically depend on a number of factors, including the exact drug, the dosage, the way of administration (e.g. local or systemic), and the specific behavioral or neural outcome measure under study.

Based on the reported literature, it might be concluded that the single dose of 2 mg haloperidol used in the current study could in principle have affected DA function in the PFC via direct or indirect mechanisms, albeit presumably to a lesser extent than in the striatum. Yet, various factors may have influenced the individual expression of such prefrontal drug actions, thereby contributing to the complex pattern of behavioral drug effects observed on the subject level. For instance, individual baseline differences in the (prefrontal) D2 receptor expression (see e.g. Cohen et al., 2007; Eisenegger et al., 2014; Kirsch et al., 2006) or in the functional connectivity within ascending DA pathways (see Gass et al., 2013) may have influenced the extent to which haloperidol affected prefrontal DA function in different subjects. While it could be speculated that the observed subject-level haloperidol effects on directed exploration (i.e. the φ parameter) may actually reflect prefrontal drug actions to some part, these effects might likewise result from pre- or postsynaptic drug actions on the striatal level and a corresponding shift in the relative striatal/prefrontal DA balance (see 6.2.1.2). Hence, it is not directly inferable from the behavioral results (and without clear neural drug effects, see 6.3.2) if and how haloperidol actually modulated prefrontal DA function in the current study. Future research is necessary to further investigate the mechanisms by which D2 drugs influence specific aspects of DA function in different brain regions and how these drug actions affect explore/exploit behavior dependent on individual variations in the D2 receptor system.

6.3 fMRI results

6.3.1 Neural signatures of exploration and exploitation

First, the neural signatures of exploratory and exploitative choices were examined across all subjects and drug conditions. Consistent with previous research, it was found that the pattern of brain activity markedly differed between both types of choices.

On the one hand, exploratory choices were associated with higher activity in the FPC, IPS, dACC, and AI, replicating the results of previous human fMRI studies (Addicott et al., 2014; Daw et al., 2006; Laureiro-Martínez et al., 2014, 2015). Note that while the dACC and AI were not reported by Daw et al. (2006), this might be explained by the much smaller sample size and lower power of this study compared to the later fMRI studies. A detailed elaboration on the functional roles of these brain regions for exploratory decision making was already provided in the introduction (see 1.1.4). To briefly summarize, a growing body of evidence supports the view that the FPC and IPS are part of a

frontoparietal control network underlying exploratory decision making. In this network, the FPC may track information relevant for exploratory decisions, such as the expected reward and uncertainty of unchosen choice options, and trigger a behavioral switch from an exploitative to an exploratory mode whenever the accumulated evidence supports such a decision (see Badre et al., 2012; Boorman et al., 2009, 2011; Cavanagh et al., 2012). The IPS, in contrast, may act as an interface between these frontal areas and motor output areas, in which behavioral responses like button presses are initiated to implement exploratory actions (see Boorman et al., 2009; Daw et al., 2006; Laureiro-Martínez et al., 2015). The dACC and AI, on the other hand, are though to form a salience network involved in detecting and orienting towards salient stimuli (Menon, 2015; Uddin, 2015), which might also subserve attentional and behavioral switching from an exploitative to an exploratory mode. Furthermore, both regions have been widely implicated in risky decision making and in mediating the effect of emotional arousal on decision making, two aspects which are further discussed below (see 6.3.3). Aside from these regions, also the bilateral thalamus, cerebellum, and supplementary motor area showed greater activation during exploration compared to exploitation, mostly consistent with the findings of previous fMRI studies (see 1.1.4; Addicott et al., 2014; Daw et al., 2006, supplement; Laureiro-Martínez et al., 2014, 2015).

Regarding the neural signatures of exploitative choices, previous findings have been more mixed (see 1.1.4), but are still largely consistent with the results of this study. First, exploitative choices were associated with greater activation in the vmPFC and OFC, replicating the findings of previous work (Laureiro-Martínez et al., 2014, 2015). While Daw et al. (2006) did not find any brain regions with significantly higher activity for exploitative compared to exploratory choices, they showed that activity in both the OFC and vmPFC correlated with variables underlying value-based exploitative decision making. More specifically, they found activity in the medial OFC to correlate with the magnitude of the obtained reward, and activity in the vmPFC and medial/lateral OFC to correlate with the choice probability of the chosen option, the latter reflecting the expected reward of that option relative to the unchosen ones (see also Boorman et al., 2009). Hence, the observation of a value-related activity in the OFC and vmPFC during exploitative behavior agrees well with evidence from numerous human neuroimaging studies, which implicate these regions in encoding reward and guiding value-based decision making, as already reviewed in the introduction (see 1.1.4; e.g. Bartra et al., 2013; Grabenhorst & Rolls, 2011; Kringelbach & Rolls, 2004; O'Doherty, 2004, 2011; Peters & Büchel, 2010).

In addition, greater activation during exploitative trials was also observed in the PCC, angular gyrus, precuneus, and hippocampus, partly replicating the results of earlier studies (Addicott et al., 2014; Laureiro-Martínez et al., 2014, 2015). Together with the medial PFC, these regions are hypothesized to form a large-scale brain system referred to as the "default mode network" (DMN; Andrews-Hanna, Smallwood, & Spreng, 2014; Buckner, Andrews-Hanna, & Schacter, 2008; Raichle et al., 2001). The DMN was shown to be active during conscious rest and mind wandering, i.e. when the individual is focused on internal (self-generated) thoughts rather than on the external world or on goal-oriented

tasks. In contrast, activity in the DMN decreases when the brain is engaged in attention-demanding cognitive tasks, i.e. when cognitive resources are needed for efficient task performance (Buckner et al., 2008; Mazoyer et al., 2001; McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003). Hence, the observed activity pattern for exploitative choices might indicate that subjects were engaged in task-independent thoughts (i.e. mind-wandering) during periods of exploitation, in which the attentional demand of the bandit task should have been relatively low. In any case, it can be assumed that attentional task demands during periods of exploitation, in which the same choice was repeated every few seconds, were considerably lower than during phases of exploration, in which different choice options need to be sampled and evaluated to decide which option to choose next. Accordingly, this lower task demand during exploitation compared to exploration might explain why exploitative trials were associated with higher activity in regions of the DMN. Interestingly, the salience network (i.e. dACC and AI) has been proposed to play a key role in switching brain activity from introspective functions of the DMN to externally focused, task-based functions (Bressler & Menon, 2010; Menon, 2015), consistent with the finding that regions of this salience network showed higher activity during exploratory trials (see above). Aside from this interpretation in terms of lower task demands, certain regions of the DMN may also serve other, more specific functions during exploration. The angular gyrus, for example, has been shown to play a crucial role in number comparisons (Göbel, Walsh, & Rushworth, 2001), and its activation during exploitation may also be related to the numerical monitoring of point (reward) values (see Addicott et al., 2014). The PCC is considered to be part of the brain's valuation system and might hence also be involved in encoding reward-related information during exploitation (Bartra et al., 2013; Grueschow, Polania, Hare, & Ruff, 2015; Hayden, Nair, McCoy, & Platt, 2008; Lebreton, Jorge, Michel, Thirion, & Pessiglione, 2009). Moreover, increased hippocampal activity during exploitation may reflect processes of episodic memory retrieval involved in reward-based decision making (see Bornstein, Khaw, Shohamy, & Daw, 2017). More specifically, it has been proposed that subjects may use memories for individual instances of past choices (i.e. episodic samples) to predict the outcome of the current decision, and that the hippocampus could play a crucial role in this process (Bornstein et al., 2017; Bornstein & Norman, 2017; see also Shadlen & Shohamy, 2016; Wimmer & Shohamy, 2012). Finally, exploitative trials were also associated with higher activity in the bilateral temporal lobes, including the middle and superior temporal gyri, largely replicating the results of Addicott et al. (2014) and Laureiro-Martínez et al. (2014, 2015). While these temporal lobe regions are mostly known for auditory, language, and semantic processing (Price, 2010), they were also shown to form a subcomponent of the DMN involved in the retrieval of (social) semantic and conceptual knowledge (Andrews-Hanna et al., 2014), which might explain their activation during exploitative trials (see above). Finally, aside from the neural signatures of exploratory and exploitative choices, the reward prediction error signal was found to positively correlate with activity in the bilateral ventral striatum, consistent with numerous previous studies (e.g. Abler, Walter, Erk, Kammerer, & Spitzer, 2006; Bray & O'Doherty, 2007; Gläscher, Daw, Dayan, & O'Doherty, 2010; Hare,

O'Doherty, Camerer, Schultz, & Rangel, 2008; O'Doherty et al., 2003, 2004; see also review by Dreher, 2013).

In a next step, it was analyzed whether the subdivision of exploratory trials into directed explorations (following the highest exploration bonus) and random explorations (not following the highest exploration bonus) revealed different neural substrates for both types of exploration. Across all subjects and drug conditions, it was found that random exploration was associated with higher activity in a small region of the right FPC compared to directed exploration. Yet, after accounting for the unequal number of trials in both exploration conditions, directed and random exploration showed no longer a significant difference in their neural activity patterns. At first glance, these findings are unexpected, given that previous studies have associated FPC function with uncertainty-driven exploration (see 1.1.4; e.g. Badre et al., 2012; Cavanagh et al., 2012; Zajkowski et al., 2017), whereas random exploration has been hypothesized to rely on subcortical substrates (e.g. Aston-Jones & Cohen, 2005; Humphries et al., 2012; Ishii et al., 2002; Mandali et al., 2015). According to these studies, an increased FPC activity should have been observed during directed compared to random exploration, not vice versa. However, recent evidence suggests that also random exploration might be driven by uncertainty (Gershman, 2018) and could hence also rely on prefrontal structures to track this uncertainty and to control the level of choice stochasticity based on this metric. In addition, Mansouri et al. (2017) recently proposed a functional model of the human FPC, in which distinct subregions of the FPC play different functional roles in exploratory behavior. More specifically, they suggest that the lateral FPC may be involved in directed exploration, which entails an online tracking of relevant choice alternatives in order to potentially re-engage one of these alternatives as replacement for the currently exploited strategy. In contrast, the medial FPC may be involved in undirected (random) exploration, for which only the ongoing strategy is monitored to potentially redistribute cognitive resources away from this strategy when it is deemed irrelevant. Hence, the above finding of a small FPC subregion with higher activity during random than directed exploration would therefore be consistent with this functional model. Finally, it should be noted that a clear distinction of choices into random and directed explorations is difficult for several reasons. First, the cognitive model applied here (Bayes-SM+EP) nests the exploration bonus within the softmax function, i.e. it adds randomness also to uncertainty-driven choice behavior, making it difficult to conceptually separate both exploration strategies. Also, since the exploration bonus influences choice behavior parametrically, a choice can still be uncertainty-driven to some extent, even if it does not follow the highest exploration bonus. Furthermore, the aspect that also random exploration may be driven by uncertainty (see above; Gershman, 2018) is not captured by the cognitive model used in this study, which might therefore not describe subjects' exploration strategies appropriately. Hence, the failure to observe distinct neural correlates for random and directed exploration might also be attributable to these factors, and further studies on this topic should use behavioral paradigms and cognitive models that allow for a better distinction between both exploration strategies, such as the horizon task (Wilson et al., 2014).

Finally, a third analysis was performed to examine the neural correlates of two model-based quantities that are tightly involved in explore/exploit behavior: the expected reward ($\hat{\mu}^{pre}$) and uncertainty ($\hat{\sigma}^{pre}$) of the chosen bandit. It was found that the neural correlates of the expected reward and uncertainty largely overlapped with the brain activity patterns for exploitative and exploratory choices, respectively. These findings are not surprising, given the high positive correlation between expected reward and exploitation on the one hand, and between uncertainty and exploration on the other hand. Still, these correlations should be taken into account when interpreting neural signatures of exploration and exploitation. The first correlation reflects the fact that exploitation is most commonly, and also here, defined as choosing the option with the highest expected reward (for alternative definitions see Mehlhorn et al., 2015). Accordingly, the neural signatures of exploitation should overlap with brain regions involved in encoding expected rewards, consistent with the finding that the vmPFC and OFC were more active during exploitative trails (see above). The second correlation results from the fact that subjects usually select the option with the highest expected reward more often, and hence have more uncertainty about the alternative options, i.e. when they explore. Accordingly, the neural signatures of exploration should overlap with brain regions involved in encoding reward uncertainty, for which the FPC, but also the AI and dACC are likely candidates (see 1.1.4; e.g. Badre et al., 2012; Cavanagh et al., 2012; Christopoulos et al., 2009; Dreher, 2013; Singer et al., 2009). In fact, this second aspect is also tightly related to the criticism that reward and uncertainty (information) are confounded in the bandit task (see Wilson et al., 2014), and alternative paradigms have been proposed to remove this confoundation (see 1.1.2). In the horizon task (Wilson et al., 2014), for example, each block starts with several forced choice trials, by which the level of information (uncertainty) can be manipulated independently of the expected reward. In the "observe or bet" task (Tversky & Edwards, 1966), on the other hand, reward and information are entirely dissociated, as exploit (bet) trials only yield reward but no information, and explore (observe) trials only information but no reward. Interestingly, the neural signatures of explore and exploit trials in this task only include a subset of the activated brain regions observed in the current study. In particular, Blanchard and Gershman (2018) found greater activation during exploratory trials only in the AI, dACC, and thalamus (but not in the FPC or IPS), and during exploitative trials only in the vmPFC. Hence, these mixed findings demonstrate that the neural correlates of explore/exploit decisions clearly depend on the behavioral paradigm in which they are studied and how it conceptualizes exploration and exploitation in terms of reward and information. While the classical bandit paradigm has been criticized for its reward/information confound, it might actually represent real-world explore/exploit problems quite appropriately, which often come with a coupling of reward and information (i.e. higher expected reward and lower uncertainty during exploitation compared to exploration; see examples in 1.1.1). Still, alternative paradigms, which entirely dissociate reward and information or remove their confoundation, might also be needed to further disentangle the different aspects underlying explore/exploit decisions and their distinct neural correlates.

6.3.2 Brain activity differences between drug conditions (planned fMRI analysis)

After examining the neural signatures of explore/exploit decisions across all drug conditions, it was tested whether these signatures were modulated by the DA drugs. However, no significant drug effects were found on the brain activity patterns for exploratory choices (overall, directed, or random), nor on the neural correlates of exploitative choices or the reward prediction error. In the following, these findings are first discussed with respect to the L-dopa condition, before the haloperidol condition will be regarded.

For the L-dopa condition, the null results on the neural level are especially surprising, given the finding of a clear behavioral L-dopa effect on directed exploration across subjects (see 5.2.2 to 5.2.4). According to the initial hypothesis, DA drug effects on exploratory behavior should be associated with drug-induced changes in the activity of brain regions implicated in exploratory choices, foremost the FPC and IPS, for which no evidence was found in the fMRI data. Yet, deviating from this hypothesis, the observed L-dopa effect on explore/exploit behavior could also be assumed to rely on an enhanced phasic DA release and prediction error signaling in the striatum, as discussed above (see 6.2.1.1). Accordingly, L-dopa would be expected to increase the magnitude of the striatal reward prediction error signal, as previously shown by Pessiglione et al. (2006). However, this effect could not be replicated in the current study. A number of factors might have contributed to the failure to detect any significant L-dopa effects on the neural correlates of explore/exploit decisions or the reward prediction error. First, this failure may simply be due to a lack of statistical power provided by the sample size of 31 subjects (see Button et al., 2013; Szucs & Ioannidis, 2017; Turner, Paul, Miller, & Barbey, 2018). In fact, recent empirical work suggests that in a typical fMRI approach testing for within-group differences in brain activation, sample sizes of at least 40 should be acquired to reliably detect regions with high effect sizes, while sample sizes closer to 80 are needed to reliably detect regions with medium-sized effects (Geuter, Qi, Welsh, Wager, & Lindquist, 2018; see also Turner et al., 2018). Furthermore, the timing of the drug administration might also be a crucial factor to consider, given the narrow pharmacokinetic time window of L-dopa (see 1.2.3). Previous studies have shown that the mean time point at which L-dopa reaches peak plasma concentration (t_{max}) usually lies between 30 and 60 min in healthy human subjects (e.g. Baruzzi et al., 1987; Crevoisier, Hoevels, Zürcher, & Da Prada, 1987; Iwaki et al., 2015; Keller et al., 2011; Nyholm et al., 2012; see also reviews by Contin & Martinelli, 2010; Hälbig & Koller, 2007; Khor & Hsu, 2007). Hence, the time schedule of the current experiment was adjusted to this short t_{max}: The bandit task started 30 min and ended 80 min after L-dopa administration, such that peak plasma concentrations were approximately reached halfway through the task. However, it is also conceivable that L-dopa effects on phasic DA activity might rather peak with some delay to the t_{max}, considering that L-dopa needs to pass the blood-brain barrier (by active transport), be converted to DA and packaged into synaptic vesicles to contribute to phasic DA signaling. Also, the t_{max} parameter is usually measured after over-night fasting, whereas subjects in the current study did not fast before the experiment to avoid fasting-related effects on explore/exploit behavior.

This aspect might also have reduced or delayed L-dopa effects in this study, as it was shown that food intake can significantly influence L-dopa pharmacokinetics due to slower gastric emptying and increased competition (between L-dopa and dietary amino acids) for active transport across the intestine and blood-brain barrier (Baruzzi et al., 1987; Contin & Martinelli, 2010; Nutt, Woodward, Hammerstad, Carter, & Anderson, 1984; Wang et al., 2017). Notably, the study of Pessiglione et al. (2006), in which L-dopa was found to significantly increase striatal prediction error signaling, followed a different time schedule for the drug administration. There, the behavioral task only started one hour after L-dopa administration, which could have been more suitable to capture the time interval in which L-dopa exerts its maximal neural and behavioral effects. Finally, it should be considered that the BOLD signal does not directly measure DA release, and that the precise physiological relationship between DA release and BOLD signal are currently unknown (see Brocka et al., 2018; Knutson & Gibbs, 2007). Based on evidence from pharmacological MRI studies, it has been suggested that striatal DA release may increase the BOLD signal via a D1-dependent mechanism, according to which D1 receptor activation changes the postsynaptic membrane potential and engages metabolic processes, which in turn lead to increased oxygen utilization followed by an elevated local BOLD response (Knutson & Gibbs, 2007). However, a recent optogenetic study in rats suggests that canonical BOLD responses in the reward system may actually mainly represent the activity of non-dopaminergic neurons, such as glutamatergic projecting neurons (Brocka et al., 2018). Moreover, the authors conclude from their findings that mesolimbic DA release and concurrent BOLD signal changes in regions of the reward network may not even be causally related. Given these findings, it cannot be directly inferred from the absence of a significant L-dopa effect on the BOLD signal that DA release was not affected by the drug. It is also conceivable that L-dopa might have enhanced striatal DA release to some extent without actually triggering a (detectable) BOLD signal change.

For the haloperidol condition, the null findings on the neural level are less surprising, given the lack of a clear behavioral haloperidol effect across subjects (see 5.2.2 and 5.2.3). As discussed before, it might be assumed that the low dose of haloperidol used in this study exerted a mixture of both presynaptic (DA-stimulating) and postsynaptic (DA-antagonizing) effects across subjects, presumably explaining why no overall haloperidol effects were found on the behavioral and neural level. Similarly, another human fMRI study (Pine et al., 2010) also failed to observe any significant low dose (1.5 mg) haloperidol effect on reward-related striatal activity or choice behavior across subjects. The authors discuss a number of factors which might have contributed to these null findings, including potential DA-stimulating effects due to the low dosage and the drug's relatively unspecific and widespread pharmacological effects (see 1.2.3). On the other hand, these null findings contrast with the results of Pessiglione et al. (2006), who showed that haloperidol (1 mg) reduces the magnitude of the striatal reward prediction error signal and exploitative behavior relative to L-dopa. Yet, it cannot be inferred from these results to what extent haloperidol alone affected striatal signaling in their study, as they report haloperidol effects only in relation to the L-dopa condition (since the placebo condition was not

double-blinded). From their behavioral data, it might be assumed that the reported effect was mainly driven by the drug actions of L-dopa and not haloperidol (see 6.2.2.1). Hence, it cannot be excluded that the inconsistent haloperidol findings between their study and the current one are actually attributable to differences in the L-dopa effects between both studies, for which potential reasons have been discussed above (e.g. the timing of drug administration). In addition, the failure to find a significant haloperidol effect on the neural level may also be due to a lack of statistical power, especially since the administered haloperidol dose was relatively low and might only have produced subtle changes in DA signaling. Also, these subtle changes in DA signaling could have been too weak to actually trigger a (detectable) BOLD signal change at all, as already discussed above. Future studies should consider using higher doses of haloperidol to achieve more consistent antidopaminergic effects from postsynaptic D2 receptor blockade across subjects, or other DA antagonists with a lower side effect profile.

6.3.3 Drug effects on uncertainty-related brain activity (exploratory fMRI analysis)

Aside from the points discussed so far, the null findings on the neural level might also indicate that the planned fMRI analysis simply failed to capture the specific aspect of brain function modulated by the drugs. For example, it is possible that L-dopa reduced directed exploration not by affecting the neural signatures of explore/exploit decisions or the reward prediction error, but instead by modulating some other aspect of DA-dependent brain function involved in the explore/exploit trade-off. Specifically, it was hypothesized that L-dopa might affect the neural correlates involved in behavioral switching from exploitation to exploration in response to accumulating uncertainty. Thereby, L-dopa might delay the time point at which uncertainty-driven exploration is triggered, resulting in fewer directed explorations over time. This alternative hypothesis was tested with an additional model-based fMRI analysis, in which trial-by-trial estimates for the overall uncertainty, quantified by the summed standard deviation over all bandits, were used as a parametric regressor in the GLM.

First, it was found that activity in the bilateral insula and dACC positively correlated with the overall uncertainty in the placebo condition, suggesting that these regions may either track the overall uncertainty directly or encode an affective or motivational state that increases with accumulating uncertainty. Either way, these signatures could be involved in triggering exploratory behavior under conditions of high overall uncertainty, potentially by facilitating attentional and behavioral switching between the currently exploited option and salient, more uncertain choice alternatives (see 1.1.4). Indeed, previous research on the role of the insula and ACC in human decision making supports these assumptions. For instance, numerous studies have found greater activation in these regions during risky decision making, i.e. decision making with an uncertain outcome, and have implicated both regions in encoding outcome uncertainty or risk (Christopoulos et al., 2009; Critchley et al., 2001; Fitzgerald et al., 2010; Fukunaga et al., 2018; Huettel et al., 2005; Preuschoff et al., 2006, 2008; Rudorf et al., 2012; see also reviews by Bach & Dolan, 2012; Dreher, 2013; Singer et al., 2009). Furthermore,

the insula is considered to play a key role for integrating interoceptive signals about bodily states into conscious feelings that influence decision making under risk and uncertainty (Craig, 2002, 2009; Critchley, 2005; Critchley & Harrison, 2013; Naqvi & Bechara, 2009; Singer et al., 2009). In particular, evidence from a human fMRI study (Xue et al., 2010) suggests that the insula might signal the urge for taking a risk, consistent with its critical role in signaling other feelings of urgency (Brody et al., 2002; Garavan et al., 2000; Lerner et al., 2009; Pelchat, Johnson, Chan, Valdez, & Ragland, 2004; see also review by Naqvi & Bechara, 2009). Specifically, this study found that decision making after refraining from a previous risk (i.e. gamble) was more risky and more likely to activate the insula and ACC. Moreover, this increase in insular activity was positively correlated with the increase in risk taking after refraining from a risk both within and across subjects, and also with an individual's personality trait of urgency. With respect to the current study, these findings might implicate that foregoing a previous chance to explore could increase the urge to explore and the extent of exploratory decisions in the subsequent trial, as mediated by the insula and ACC. As further discussed by Xue et al. (2010), the insula and ACC may both be part of an "integral neural network that underlies the effect of emotional arousal on risky decision-making" (p. 715). Accordingly, the co-activation of both regions has been observed in various emotional tasks (see Craig, 2009), and it has been proposed that while the insula might mediate the feeling of an emotion, the ACC might mediate the motivation associated with an emotion and hence be directly involved in the initiation of behaviors (Craig, 2002, 2009; see also Hampton & O'Doherty, 2007). In addition, the ACC has been implicated in signaling the salience of each new piece of information for predicting future outcomes (Behrens et al., 2007; Rushworth & Behrens, 2008) and may hence be involved in guiding attention and actions towards choice options that are especially uncertain or informative. Consistent with this idea, a human fMRI study (Christopoulos et al., 2009) found that activity in the dACC not only increased with risk (i.e. outcome uncertainty), but was directly linked to subjects' choice behavior in a monetary gambling task, as it positively predicted the probability of a risky choice. These findings also fit well to a large body of research implicating the ACC in cognitive control processes (Kerns et al., 2004; Niendam et al., 2012; Ochsner & Gross, 2005; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004) and in the underlying cost-benefit evaluations that determine which action to select next and how much control to allocate to this action (Shenhav, Botvinick, & Cohen, 2013; Shenhav, Cohen, & Botvinick, 2016). Furthermore, the ACC is also considered to play a crucial role in monitoring response conflict – which should increase with the overall uncertainty – and in triggering attentional and behavioral changes which serve to reduce conflict in subsequent performance (Botvinick, Cohen, & Carter, 2004; Kerns et al., 2004; van Veen & Carter, 2002). Finally, both the ACC and insula project to the striatum (Chikama, McFarland, Amaral, & Haber, 1997; Haber & Knutson, 2010; Kunishio & Haber, 1994) and have been suggested to directly modulate striatal reward signals and reward-related behavior (Behrens et al., 2007; Botvinick, Huffstetler, & McGuire, 2009; Jones, Minati, Harrison, Ward, & Critchley, 2011; Shenhav et al., 2013, 2016; Walton, Kennerley, Bannerman, Phillips, & Rushworth, 2006; see also Elston & Bilkey, 2017). For instance, evidence from a human fMRI study (Jones et al., 2011) suggests that feelings of urgency, as

represented in the insula, may directly interfere with striatal representations of expected reward to influence risk/reward decision making. Accordingly, it might be speculated that uncertainty-related activity in the insula and ACC, as observed in the current study, could directly interfere with the subcortical DA substrates of value-driven choice behavior to facilitate switching between exploitation and exploration. Taken together, these findings largely support the assumption that both the insula and dACC are tightly involved in triggering exploration under circumstances of high overall uncertainty. The insula might thereby signal an urge to explore, which grows with accumulating uncertainty and encourages subsequent exploration, whereas the dACC might subserve cognitive control processes that guide attention and behavioral responses towards salient, uncertain choice alternatives.

Importantly, it was also found that this uncertainty-related activity in the insula and dACC was reduced under L-dopa compared to placebo, potentially explaining why subjects showed less directed exploration under L-dopa. More specifically, it might be speculated from these findings that subjects in the L-dopa condition felt a lower urge to explore under conditions of high overall uncertainty and were less drawn towards uncertain choice alternatives, hence they stayed longer at the currently exploited option. Moreover, it was found that uncertainty-related activity in the bilateral insula was also reduced under haloperidol compared to placebo. However, these haloperidol effects on insular activity were not associated with a corresponding shift in the explore/exploit trade-off on the behavioral level across subjects (see 6.2.2). Also, an additional fMRI regression analysis found no association between the subject-specific L-dopa or haloperidol effects on uncertainty-related brain activity (i.e. the BOLD correlate of overall uncertainty) and the subject-specific drug effects on directed exploration (i.e. the φ parameter). Hence, it remains unknown how exactly these DA drug effects on the neural level correspond to behavioral changes in the explore/exploit trade-off. Also, it should be noted that both drugs were found to affect activity in partly different subregions of the insular cortex (i.e. left/right, anterior/posterior), whereas it is not clear what functional roles these different subregions may play in the neural response to uncertainty and the regulation of explore/exploit behavior. Finally, it remains unknown by what mechanism L-dopa and haloperidol might have affected these signals in the insula and ACC. Previous research has shown that both regions receive DA projections from the midbrain (Berger, Gaspar, & Verney, 1991; Narita et al., 2010; Ohara et al., 2003), express D1 and D2 receptors (Gaspar, Bloch, & Moine, 1995; Hurd et al., 2001; Richfield, Young, & Penney, 1989), and that their function is modulated by DA and DA drugs (e.g. Burkey, Carstens, & Jasmin, 1999; Coffeen et al., 2008, 2010; López-Avila, Coffeen, Ortega-Legaspi, del Angel, & Pellicer, 2004; Narita et al., 2010; Schweimer & Hauber, 2006; see also reviews by Assadi, Yücel, & Pantelis, 2009; Coffeen, Ortega-Legaspi, & Pellicer, 2012; Gogolla, 2017). Hence, it is possible that L-dopa and haloperidol *directly* modulated DA transmission – and thereby the BOLD signal – in these cortical regions. However, a drug-induced increase in DA release by L-dopa or haloperidol (assuming presynaptic haloperidol effects) should have increased rather than decreased the BOLD signal in these regions (see Knutson & Gibbs, 2007), conflicting with the reported findings. Moreover, human studies

have shown that the expression of D1 and D2 receptors is much higher in the striatum than in the insula and ACC (Hall et al., 1994; Hurd et al., 2001), and both L-dopa and haloperidol are considered to primarily exert their effects on the striatal level (as discussed above, see 6.2). Hence, it seems more likely that these drugs affected uncertainty-related activity in the insula and ACC *indirectly* by modulating DA transmission on the striatal level. More specifically, it might be assumed that information about reward uncertainty is initially encoded on the striatal level and then transmitted to cortical structures to be integrated with other decision parameters for guiding behavior (see Haber & Knutson, 2010; Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Rushworth & Behrens, 2008; Shenhav et al., 2013). Indeed, a number of studies in humans and nonhuman primates have provided evidence that the striatal DA system codes both the expected value and uncertainty (i.e. variance) of reward via spatially and temporally distinct signals (Dreher, 2013; Dreher, Kohn, & Berman, 2006; Fiorillo, Tobler, & Schultz, 2003; Preuschoff et al., 2006; Schultz et al., 2008; see also Linnet et al., 2012; Rudorf et al., 2012). For instance, a human fMRI study (Preuschoff et al., 2006) has shown that during reward anticipation in a monetary gambling task, initial BOLD activation in the striatum and putamen correlates positively with expected reward, whereas delayed BOLD activation in the striatum and midbrain shows an inverted-U relationship with reward uncertainty. The notion that the midbrain and striatal DA system codes reward uncertainty according to an inverted-U relationship is also supported by evidence from single cell recordings in monkeys (Fiorillo et al., 2003; Schultz et al., 2008) and human PET imaging (Linnet et al., 2012). Given these findings, it might be interesting to reexamine the fMRI data of the current study, as they could potentially reveal striatal BOLD correlates of reward uncertainty and a modulation of these correlates by the DA drugs.

Taken together, while this exploratory fMRI analysis only revealed weak DA drug effects on uncertainty-related brain activity, these effects still point to an interesting hypothesis about how L-dopa could have affected explore/exploit behavior. To empirically validate these ideas, future studies should more closely examine the role of the insula and ACC in triggering exploration in response to accumulating uncertainty and further investigate how DA might be involved in this process.

6.4 Inverted-U analysis

One additional aim of this study was to test whether DA drug effects on explore/exploit behavior were modulated by the individual DA baseline (indexed by the sEBR and WMC), as predicted by the inverted-U hypothesis of DA. However, the current study found no evidence for such a relationship between the DA baseline measures and DA drug effects on explore/exploit behavior. Still, one should be cautious to interpret this finding as clear evidence against the inverted-U hypothesis in general for several reasons.

First, this finding contrasts with a large body of research supporting the inverted-U hypothesis (see 1.2.4; reviewed by Cools & D'Esposito, 2011; Floresco, 2013), some of these studies even assessing

baseline DA functions more directly with PET imaging (Cools et al., 2008; Landau et al., 2009). Moreover, it should also be noted that despite this evidence, the inverted-U hypothesis of DA remains relatively vague in its predictions and is therefore rather difficult to test and falsify. For example, it remains unclear how to construe the exact shape and turning point (optimum) of the inverted-U function, especially since it has been suggested that these features may vary between different tasks, cognitive functions, and individual subjects (see Cools et al., 2009; Cools & D'Esposito, 2011; Fallon et al., 2015; Wiegand et al., 2016). Furthermore, it is not clear to which specific aspect(s) of DA function and which cognitive domains this hypothesis actually applies. For example, animal studies suggest that the inverted-U function specifically describes the relationship between prefrontal D1 receptor activity and working memory performance, whereas the relation between other DA aspects and cognitive functions may follow different functions (see Floresco, 2013; Floresco & Magyar, 2006). Hence, it cannot be unequivocally concluded from the data of the current study if they generally speak against the inverted-U hypothesis, or if this hypothesis simply not applies to the specific aspect of DA function and behavior under study.

Aside from these points, it should also be considered that the current study has a number of limitations, which might have contributed to the failure to observe an inverted-U effect in the data. First, the sample size of 31 subjects may have been too small to provide enough power for the statistical tests performed on account of the inverted-U hypothesis, i.e. the test for an inverse quadratic relationship in the baseline (placebo and pilot) data and the test for a group difference in DA drug effects between subjects with low vs. high DA baseline measures. To detect these inverted-U effects, it is critical to have a subject sample with sufficient variability in the DA baseline measures, ideally also including extreme values at both ends of the curve, which might require a much larger sample size than used here. In relation to this point, it has been argued (Slagter et al., 2012) that healthy subjects may display only a relatively restricted range in baseline DA levels during resting conditions, making it more difficult to observe inverted-U effects in these samples. A further limitation of this study relates to the observation that the distribution of sEBR values was strongly left-skewed across subjects, with only few high-sEBR subjects in the sample. This aspect proves particularly problematic when splitting the sample into low- and high-sEBR groups to test for DA baseline effects, which also relates to the aforementioned problem of how to define the turning point of the inverted-U curve: On the one hand, one might define the turning point as the median value of the sample, which is the typical approach when testing for DA baseline effects (see review by Jongkees & Colzato, 2016). While median splitting provides similar group sizes, it also entails the problem that the position of the turning point heavily depends on the distribution of baseline values in the sample and may hence be ill-defined if this distribution is strongly skewed. In the current study, for example, the sEBR median value of 11.2 was much closer to the smallest (5.4) than to the largest (38.4) observed sEBR value in the sample. Alternatively, one might define the turning point as the midpoint between the smallest and largest observed sEBR value and split the sample at this point (here: 21.9). However, when using

this approach in the current study, only three subjects remained in the high-sEBR group, also providing no reliable basis for statistical inference. A larger sample size or the pre-screening of subjects with respect to their DA baseline values (e.g. sEBR or WMC) may circumvent such problems by providing a testing sample with a more even and wide-ranged distribution of these values to facilitate testing for inverted-U effects.

Aside from these limitations, the failure to observe an inverted-U effect in the current study might also be explained in terms of poor DA proxy measures. While the sEBR has been extensively used as a proxy for DA function in animals and humans, many of these studies have also yielded conflicting results (see 1.2.5; reviewed by Jongkees & Colzato, 2016). For example, some studies have found that the sEBR is sensitive to pharmacological modulation of the DA system in healthy humans (e.g. Blin, Masson, Azulay, Fondarai, & Serratrice, 1990; Cavanagh et al., 2014), whereas other studies have found no such effects (e.g. Ebert et al., 1996; Mohr et al., 2005; van der Post, de Waal, de Kam, Cohen, & van Gerven, 2004; see also Jongkees & Colzato, 2016). In accordance with the latter, the present study has found no DA drug effects on the sEBR with either L-dopa or haloperidol. Furthermore, recent PET experiments in humans have raised some doubt about the validity of the sEBR as a (positive) predictor of DA, as they have found either no or even a negative relationship between the sEBR and different aspects of central DA function (Dang et al., 2017; Sescousse et al., 2018). For the WMC, the available evidence is overall more limited than for the sEBR, as fewer studies have used this measure as a DA proxy (see 1.2.5). While some studies have found that the working memory span predicts individual differences in DA drug effects on cognitive performance, they also yielded opposing results with respect to the direction of these effects. For example, while two studies found that the DA agonist bromocriptine improves working memory performance in low-span subjects, but impairs it in high-span subjects (Gibbs & D'Esposito, 2005; Kimberg et al., 1997), two other studies found the opposite effect with the DA agonist pergolide (Gibbs & D'Esposito, 2006; Kimberg & D'Esposito, 2003). To reconcile these mixed findings, it has been proposed that baseline effects of the working memory span may depend on a number of factors, including the specific DA function targeted by the drug (D1 vs. D2), the site of modulation (striatal vs. frontal), and the behavioral outcome measure under study (Cools & D'Esposito, 2011; see also Fallon et al., 2015; Kimberg & D'Esposito, 2003). Yet, the proposed complexity of interactions between these different factors further complicates empirical testing of inverted-U effects with the WMC. Also, direct evidence for the assumption that the WMC predicts central DA function in humans remains relatively sparse to date (Cools et al., 2008; Landau et al., 2009). Moreover, for both proxies, WMC and sEBR, it is not yet clear which specific aspect(s) of DA function they predict and why, i.e. by what neural mechanism. Given this uncertainty, it is also conceivable that these proxies reflect a specific aspect of DA function, which is however not the critical determinant for the inverted-U effect with respect to a studied behavior. For example, previous research suggests that the inverted-U curve specifically describes the relationship between *prefrontal* D1 receptor function and working memory performance (see above; e.g. Fallon et al., 2015; Floresco, 2013; Floresco &

Magyar, 2006), whereas the sEBR and WMC may rather reflect certain aspects of *striatal* DA function, such as striatal D2 receptor availability (see Groman et al., 2014; Jongkees & Colzato, 2016) and striatal DA synthesis capacity (Cools et al., 2008; Landau et al., 2009), respectively. In conclusion, it cannot be ruled out that the DA proxies used in this study either failed to validly measure baseline DA function, or specifically measured an aspect of DA function which was not predictive for the behavioral outcome measure under study.

Finally, it should be considered that the explore/exploit trade-off describes a very complex behavior, which differs from the behavioral outcome measures that have usually been used to study the inverted-U hypothesis. Previous studies have mostly examined inverted-U effects on behavioral variables for which an "optimal performance" (i.e. good or bad performance) can be more readily defined, such as accuracies and reaction times in working memory tasks (e.g. Frank & O'Reilly, 2006; Gibbs & D'Esposito, 2005, 2006; Mehta et al., 2000), or perseverative errors in set shifting tasks (e.g. Frank & O'Reilly, 2006; Kimberg et al., 1997). However, it is theoretically not clear what level of directed or random exploration may be "optimal" (see 1.1.1; Cohen et al., 2007) and how to describe explore/exploit behavior in terms of an optimum curve. Accordingly, the model-based variables used here to quantify exploratory choice behavior (β and φ) may therefore not be appropriate dependent variables in the inverted-U analysis. On the other hand, alternative (model-free) performance measures like the overall payout may be too crude to reflect a DA-specific cognitive function, explaining why also no inverted-U effects could be observed on these variables.

Taken together, although the results of the current study do not support the inverted-U hypothesis of DA, a number of limitations have been discussed that may have contributed to the observed null findings. Future studies should try to overcome these limitations, e.g. by larger sample sizes and/or direct PET assessment of baseline DA function, in order to further refine the inverted-U hypothesis and show if it also applies to (certain aspects of) explore/exploit behavior.

6.5 Cognitive model comparison

A further aim of this study was to quantitatively compare different cognitive models of learning and decision making for their predictive accuracy in the examined explore/exploit paradigm. In the following, the results of this model comparison are first discussed with respect to the applied learning rules, before the different choice rules are considered in more detail.

6.5.1 Learning rules

Two different learning rules were used to describe subjects' learning process in the bandit task: a simple reinforcement learning rule (Delta rule) and a more complex Bayesian learner rule. With regard to these learning rules, the model comparison showed that the Bayesian learner outperformed the Delta rule for each of the applied choice models. This superiority of the Bayesian learner may be

attributed to the fact that this model offers, in many ways, a more elaborate description of the learning process than the simple Delta rule and might thereby capture the complexity of human behavior more appropriately. First, while both models are routed on the same error-driven learning principle, the Bayesian learner assumes that subjects additionally track the variance (uncertainty) of their reward expectation, which is not represented in the Delta rule. In line with this assumption, several neuroimaging studies have shown that the human brain indeed tracks trial-by-trial changes in reward uncertainty as encoded in different brain regions, including the FPC, ACC, insula, and striatum (e.g. Badre et al., 2012; Cavanagh et al., 2012; Critchley et al., 2001; Dreher et al., 2006; Fukunaga et al., 2018; Huettel et al., 2005; Preuschoff et al., 2008; Rudorf et al., 2012; Yoshida & Ishii, 2006; see also reviews by Bach & Dolan, 2012; Dreher, 2013). Second, this additional tracking of uncertainties can be seen as a prerequisite for the implementation of more sophisticated choice strategies for directed exploration, as evidently employed by humans (e.g. Cogliati Dezza et al., 2017; Wilson et al., 2014). For example, if directed exploration is implemented via an exploration bonus, this bonus can be directly calculated from the trial-by-trial uncertainty estimates provided by Bayesian learning, whereas Delta learners would need to approximate this uncertainty by the use of simple heuristics (see Dayan & Sejnowski, 1996; Speekenbrink & Konstantinidis, 2015; Sutton, 1990). Indeed, previous neuroimaging studies suggest that humans use such trial-by-trial uncertainty estimates derived from Bayesian learning, as coded in frontal brain regions, to guide directed exploration (Badre et al., 2012; Cavanagh et al., 2012; see also Boorman et al., 2009). Third, the Bayesian learner implements a more efficient update algorithm than the Delta rule, since it dynamically adjusts its learning rate from trial to trial according to the current level of uncertainty. Thereby, error-driven learning is high when reward predictions are uncertain (i.e. during exploration), but decreases when predictions become more certain (i.e. during exploitation). In fact, it was shown that this update algorithm (Kalman filter) of the Bayesian learner model represents the optimal mean-tracking rule for the Gaussian type of restless bandit problem implemented in the current study (in terms of minimal mean squared errors; see Anderson & Moore, 1979; Kalman, 1960; Kalman & Bucy, 1961). Based on these points, it seems reasonable to assume that the Bayesian learner model captures the actual reward learning process as implemented in the human brain more appropriately than the simpler Delta rule, in line with the model comparison results of the present study.

On the other hand, previous model comparison studies have provided mixed evidence for the question which of these models captures human learning more accurately. For example, Speekenbrink and Konstantinidis (2015) compared the Bayesian learner and Delta rule in combination with various choice rules in a variant of the restless four-armed bandit task with both gains and losses and changing volatilities (diffusion rates). A comparison of all cognitive models across subjects showed that the performances of both learning rules were relatively matched, which a slight advantage for the Delta rule. However, when they compared model fits between subjects, they found that the best-fitting model varied between subjects and that the largest number of participants were best fitted by one of

the models incorporating the Bayesian learner. Furthermore, another human study (Payzan-LeNestour & Bossaerts, 2011) compared the model fits of a Bayesian learning rule and Delta rule (both combined with the softmax choice rule) in a restless six-armed bandit task, in which reward probabilities of all arms unexpectedly changed ("jumped") during the task. Similar to the current study, their Bayesian learner model assumes that subjects track both the expected value and uncertainty of reward on the basis of an internal representation of the underlying reward structure – also referred to as "model-based learning". The Delta rule, in contrast, makes no assumptions about the underlying reward process and learns expected rewards only on the basis of the previous reward history – also called "model-free learning". Interestingly, a model comparison across subjects showed that the Bayesian learner only fitted better when subjects were fully instructed about the underlying reward structure of the task. Without full instruction, behavior was slightly better fitted by the model-free Delta rule across subjects, whereby the best-fitting model varied between subjects. Hence, the authors concluded from these results that humans implement Bayesian learning only if enough structural information about the reward-generating process is provided. Therefore, the superiority of the Bayesian learner rule observed in the current study may (partly) reflect the fact that in this study, participants were mostly instructed about the underlying reward structure (i.e. that rewards of all options change slowly and randomly over time) and could additionally learn about the task structure in a prior training run. Also, the bandit task structure in the present study was decidedly less complicated than in the jumping six-armed bandit task used by Payzan-LeNestour and Bossaerts (2011), which might have further encouraged the utilization of a model-based learning rule. Finally, the question whether humans engage in model-based or model-free learning was also examined by Knox et al. (2012), who used a variant of the restless two-armed bandit task called the "Leapfrog task", in which both options continually alternate in their superiority by sudden jumps in their payoffs. They found that the process by which subjects update their reward expectations was better captured by a model which incorporates knowledge about the underlying reward structure (model-based learning) than by a naïve RL model which – based on the Delta rule – updates rewards only on the basis of observed outcomes (model-free learning). In conclusion, while the results of the current study suggest that human reward learning is more accurately described by the Bayesian learner than the Delta rule, previous evidence in this regard is mixed. It might be assumed from the reported findings that the best-fitting model not only varies between subjects, but also strongly depends on the kind of task, the complexity of the underlying reward structure, and subjects' prior knowledge about this structure (i.e. task instructions).

6.5.2 Choice rules

In combination with the learning rules, three different choice rules were evaluated in this study: the standard softmax rule ("SM"), a softmax with exploration bonus ("SM+E"), and a softmax with both exploration bonus and perseveration bonus ("SM+EP"). It was found that irrespective of the learning

rule and in line with the initial hypothesis, explore/exploit behavior was best described by the SM+EP model, which captures both random and directed exploration along with choice perseveration.

First, the finding that the model fit of the softmax rule is improved by inclusion of an exploration bonus parameter (φ) is consistent with previous research, showing that humans use both random and uncertainty-driven exploration (e.g. Frank et al., 2009; Gershman, 2018; Krueger et al., 2017; Payzan-LeNestour & Bossaerts, 2012; Somerville et al., 2017; Wilson et al., 2014; Zajkowski et al., 2017). As in the SM+EP model, random and directed exploration are usually implemented via two separate model parameters. On the one hand, random exploration is commonly modeled by adding noise (stochasticity) to the value-driven choice process via the softmax function (see Beeler et al., 2010; Beeler, 2012; Daw et al., 2006; Gershman, 2018; Humphries et al., 2012; Thrun, 1992). Since previous studies have already shown that this softmax function captures random exploration in humans better than the simpler ε-greedy model (see 1.1.3; Daw et al., 2006; Speekenbrink & Konstantinidis, 2015), the latter was not included in the model comparison of the current study. Uncertainty-driven exploration, on the other hand, is commonly implemented by the use of an exploration bonus, sometimes also referred to as "information bonus" or "novelty bonus" (Cogliati Dezza et al., 2017; Daw et al., 2006; Dayan & Sejnowski, 1996; Ishii et al., 2002; Kakade & Dayan, 2002; Wilson et al., 2014; Wittmann, Daw, Seymour, & Dolan, 2008). Despite the different terms, the basic idea behind this approach is to add an extra bonus for "uncertainty" or "information" to the expected reward value of each option and to select actions based on this combined value, thereby biasing choices towards more uncertain/informative options. Consistent with the results of the current study, numerous studies have found evidence for such an exploration bonus in human decision making (e.g. Badre et al., 2012; Cavanagh et al., 2012; Cogliati Dezza et al., 2017; Frank et al., 2009; Geana, Wilson, Daw, & Cohen, 2016; Gershman, 2018; Krueger et al., 2017; Wilson et al., 2014; Zajkowski et al., 2017), with only few exceptions (as discussed below; Daw et al., 2006; Speekenbrink & Konstantinidis, 2015). Notably, the current study also found evidence for an exploration bonus when reward uncertainty is not directly estimated (as in the Bayesian learner), but instead approximated by simple heuristics like the time that has passed since an option was last chosen (as in the Delta rule; see Speekenbrink & Konstantinidis, 2015; Sutton, 1990).

Interestingly, when regarding the subject-level estimates for the exploration bonus parameter φ in the current study, it was found that this parameter was actually negative for some subjects, reflecting rather a "penalty" than a "bonus" for uncertainty. In line with this finding, negative exploration (or information) bonuses have already been reported by previous studies (Cavanagh et al., 2012; Daw et al., 2006; Payzan-LeNestour & Bossaerts, 2011, 2012; Wilson et al., 2014), and different interpretations of this finding have been discussed. For example, it has been suggested that a negative exploration bonus might actually not reflect a specific explore/exploit strategy, but might rather capture choice autocorrelation, presumably resulting from perseveration (see Badre et al., 2012; Daw et al., 2006, supplement; Payzan-LeNestour & Bossaerts, 2012). Perseveration, also known as "sticky choice",

describes the behavioral tendency to repeat the same choice regardless of rewards, and a number of studies have found evidence for such a tendency in human choice behavior (Brough et al., 2008; Payzan-LeNestour & Bossaerts, 2012; Rutledge et al., 2009; Schönberg et al., 2007; Worthy et al., 2013; see also Lau & Glimcher, 2005). Crucially, if perseveration is not explicitly accounted for in the cognitive model, then it might be captured by the exploration bonus parameter as a (value-independent) choice preference for the option with the smallest uncertainty. As a result, estimates for the exploration bonus parameter will be smaller, often negative and harder to interpret (see Badre et al., 2012; Daw et al., 2006, supplement; Payzan-LeNestour & Bossaerts, 2012). To circumvent this problem, the current study has extended the softmax model with exploration bonus (SM+E; Daw et al., 2006) by an extra perseveration bonus parameter (SM+EP) and shown that the introduction of this parameter improves the model fit. Moreover, the introduction of this parameter significantly increased the φ parameter estimate across subjects and reduced the number of subjects with a negative φ parameter estimate. Hence, it can be concluded that choice perseveration is indeed captured, to some part, by the exploration bonus parameter of the SM+E model and that the SM+EP improves the model fit by disentangling these two distinct choice tendencies. Still, the introduction of a perseveration bonus parameter did not fully solve the issue, as negative φ parameter estimates were still observed in the Bayes-SM+EP model for some of the subjects. A similar result was obtained by Payzan-LeNestour and Bossaerts (2012), who modeled human choice behavior in a restless six-armed bandit task. They found that the exploration bonus parameter was negative for most of their subjects, even after accounting for perseveration in their cognitive models. Hence, these findings show that negative exploration bonuses cannot be fully explained by choice perseveration. Alternatively, it has been suggested that negative exploration bonuses may indeed be interpreted as a penalty term for uncertainty, thereby reflecting a form of "ambiguity-aversion" that discourages subjects from engaging in uncertainty-driven exploration (Badre et al., 2012; Payzan-LeNestour & Bossaerts, 2011, 2012). Payzan-LeNestour and Bossaerts (2012) further investigated this idea by fitting new and improved models to both human choice data and simulated data in the restless six-armed bandit task. Together, the results of their study support the hypothesis that uncertainty-driven exploration might actually involve a dilemma between two motives: a curiosity motive, which seeks out uncertainty to learn about novel reward opportunities, and a cautiousness motive, which avoids uncertainty for its potential dangers. Accordingly, the curiosity motive might be captured by a positive exploration bonus, and the cautiousness motive by a negative exploration bonus (i.e. penalty). Hence, the classic exploration bonus parameter, which does not distinguish between both aspects, may therefore be a "readout of the dominating motive" (Payzan-LeNestour & Bossaerts, 2012, p. 4) and range from negative to positive values as observed in the current study.

Although the modeling results of the current study are largely consistent with previous research, they contrast with the study of Daw et al. (2006), who found no evidence for an exploration bonus in human decision making. More specifically, they found that while the inclusion of an exploration bonus

parameter significantly improved the model fit for half of their subjects, the best-fitting estimate for this parameter was actually close to zero, thereby making the model equivalent to the simple softmax rule. Furthermore, when they analyzed the contribution of this parameter to exploratory behavior, they found that the majority of exploratory choices could not be explained by the bonus, suggesting that subjects predominantly relied on the softmax strategy for (random) exploration. Hence, Daw et al. (2006) concluded that there is "no evidence to justify the introduction of an extra parameter that allowed exploration to be directed towards uncertainty". The apparent inconsistency of these results with the results of the current study, and with a large body of evidence in support of the exploration bonus, might be explained by a number of factors. First, Daw et al. (2006) did no account for choice perseveration in their cognitive models, which might have contributed to the small (and often negative) φ parameter estimates observed in their study and their lack of support for the SM+E model. Second, assuming that the SM+E model captures both uncertainty-seeking and uncertainty-avoidance in the same bonus parameter (see above; Payzan-LeNestour & Bossaerts, 2012), a φ parameter estimate close to zero might not necessarily be interpreted as evidence against an exploration bonus, but could also indicate the overlay of two competing choice motives that are not separately resolved by the model. Third, it has been argued that the failure to observe exploration bonuses in the classic restless bandit task might also stem from the fact that reward and uncertainty are usually confounded in this task, since subjects tend to choose more rewarding options more often and have therefore less uncertainty about these options (see Gershman, 2018; Wilson et al., 2014). Hence, reward-driven behavior might be mistaken as uncertainty-aversion, making it more difficult to find evidence for the exploration bonus unless this confound is removed, as for example in the horizon task (see 1.1.2; Wilson et al., 2014). Yet, it should be noted that all three arguments also apply to the SM+E model used in the current study, for which φ parameter estimates were – in contrast to the results of Daw et al., (2006) – significantly different from zero for most subjects (based on their 90% HDIs, data not shown). Yet, there is another aspect which might explain the contradicting findings between the study of Daw et al. (2006) and the current study. Daw et al. (2006) obtained in their Bayes-SM+E model fit an extremely large estimate for the diffusion variance parameter ($\hat{\sigma}_d^2$), which quantifies a subject's belief about the level of reward volatility. Across subjects, the best-fitting estimate for the $\hat{\sigma}_d$ parameter was 50.9, which is several times larger than the estimates obtained in the current study when treating $\hat{\sigma}_d$ as a free parameter (between 6.4 and 7.0; see Table A1 in the appendix) and the actual diffusion variance used to generate the payoffs ($\sigma_d^2 = 2.8^2$; see 2.4). Importantly, the $\hat{\sigma}_d$ parameter largely contributes in each trial to the uncertainty estimate for each bandit ($\hat{\sigma}^{pre}$), and therefore to the exploration bonus ($\varphi\hat{\sigma}^{pre}$). Hence, a large $\hat{\sigma}_d$ value, i.e. an overestimation of reward volatility, might have two notable effects. First, it should be associated (in the joint parameter estimates) with an accordingly small φ parameter value to keep the size of the exploration bonus in a plausible range, given its direct scaling in reward units. Second, a large $\hat{\sigma}_d$ keeps uncertainty estimates for *all bandits* (regardless of choice history) constantly high, such that differences in the uncertainty and exploration bonus between chosen and unchosen bandits become negligible (as confirmed by

simulations with different $\hat{\sigma}_d$ values, data not shown). In this scenario, where diffusion variance (reward volatility) is the major source of uncertainty, the informational value of exploration will be very small since reward uncertainty cannot be substantially reduced by sampling. Both these factors might explain why φ parameter estimates were negligible in the study by Daw et al. (2006). Interestingly, a similarly large estimate for the $\hat{\sigma}_d$ parameter ($\hat{\sigma}_d$ = 52.4) in the Bayes-SM+E model was also reported by Speekenbrink and Konstantinidis (2015), who also found no evidence in support of this model compared to the simpler Bayes-SM model without the exploration bonus. The question remains, however, why $\hat{\sigma}_d$ estimates have been considerably larger in these studies than in the current one. On the one hand, it is possible that these large estimates actually captured some specific aspect of subjects' learning behavior, i.e. a general overestimation of reward volatility, which might have varied between studies due to differences in the subject sample, study design, or task instruction. On the other hand, it is also possible that the $\hat{\sigma}_d$ parameter is poorly identified in the Bayes-SM+E model (with free random walk parameters) and thereby causes unreliable estimates, especially within the MLE framework applied in both of the earlier studies. Indeed, Daw et al. (2006) reported that parameter estimates in their individual model fits pointed towards poor identifiability and often yielded extreme values, making it necessary to equate most of the free parameters (including φ) between subjects. In fact, similar problems of unreliable parameter estimates for the Bayes-SM+E model have also been encountered in the present study, despite adopting a hierarchical Bayesian modeling approach to facilitate parameter estimation. More stable parameter estimates were only obtained after constraining the free parameter space of this model by fixing the random walk parameters of the Bayesian learner rule (see Table 1), which were of secondary relevance for the current study. Although fixing these parameters also has its disadvantages (i.e. subjects' belief about the reward-generating process might not be appropriately captured), it may have contributed to more reliable and plausible estimates for the exploration bonus parameter than obtained in the earlier studies. Either way, the findings of Daw et al. (2006) and Speekenbrink and Konstantinidis (2015) should not be interpreted as clear evidence against the notion of an exploration bonus in human decision making, especially since numerous studies consistently support this idea, including the current one.

Finally, it should be noted that recent modeling research suggests that the processes underlying human explore/exploit behavior are actually much more complex than assumed by the cognitive models used in the current study. For example, it has been described (Payzan-LeNestour & Bossaerts, 2011) that there are at least four different types of uncertainty that influence human decision making, of which estimation uncertainty – modeled here as the prior standard deviation associated with each bandit – is only one. Other types of uncertainty include expected uncertainty (risk), unexpected uncertainty (e.g. sudden contingency changes), and structural uncertainty (e.g. about the reward environment), which might even be separately encoded in the human brain (see Bach & Dolan, 2012; Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005; Huettel et al., 2005; Payzan-LeNestour & Bossaerts,

2011; Schultz et al., 2008). Also, another modeling study (Krueger et al., 2017) suggests that there are actually three independent processes that drive human exploration: decision noise (random exploration), information-seeking (directed exploration), and a baseline uncertainty-seeking, which is driven by a prior that is optimistic for losses and pessimistic for gains. In other words, Krueger et al. (2017) propose that information-seeking (to reduce uncertainty) is not the same as uncertainty-seeking, and show that both processes can be distinguished in a variant of the horizon task with both gains and losses. Moreover, a very recent modeling study suggests that also random exploration might be influenced by reward uncertainty (Gershman, 2018). They report evidence for a "hybrid model" of exploration, in which uncertainty acts as both an exploration bonus and a driver for choice stochasticity, in contrast to earlier models assuming a fixed (uncertainty-independent) source of randomness like the softmax rule. Aside from these conceptual refinements of uncertainty and exploration, it was also shown that modeling of human explore/exploit behavior can be improved by accounting for individual differences in strategy (Steyvers et al., 2009). By modeling choice behavior in the stationary bandit task from a large number of subjects (n=451), they found clear evidence for individual differences in the way participants approached the explore/exploit trade-off, both in terms of which decision model they used (i.e. optimal Bayesian model vs. simpler heuristic strategies) and what the best-fitting parameter values were. Moreover, they found that both the adoption of an optimal Bayesian strategy and the ability to approximate the true underlying reward structure of the task significantly correlated with measures of general intelligence. Finally, further studies showed that human explore/exploit behavior is not only influenced by trait measures like intelligence, but also by the temporary context in which the task is performed. For example, participants were found to explore more when they perceive the task as boring (induced by experimental manipulation of the task environment; Geana et al., 2016) or when the overall reward level in the task is high (Cogliati Dezza et al., 2017). Taken together, the reported findings point to interesting ideas how modeling of explore/exploit behavior may be extended to better capture the complexity of the underlying cognitive processes and their modulation by different individual traits and situational factors.

6.6 Limitations and future directions

Aside from the limitations that have already been discussed in the preceding sections, some further limitations of this study need to be acknowledged. A first limitation relates to the fact that while the applied pharmacological fMRI approach can visualize DA drug effects on the whole brain level, it cannot determine which of these effects directly reflect local changes in DA signaling, and which reflect downstream effects that may also involve other neurotransmitter systems (see Schrantee & Reneman, 2014). As already pointed out, the fMRI BOLD signal provides an indirect index of blood oxygenation rather than a direct measure of DA activity (see 2.8.1 and 6.3.2). Hence, an observed BOLD signal change must not necessarily rely on a change in DA transmission, and a change in DA transmission must not necessarily produce a (detectable) BOLD signal change (see Brocka et al., 2018). Accordingly,

it cannot be ruled out that DA drug effects on a specific brain region (e.g. the striatum) are not actually reflected in a BOLD signal change in this region, but modulate activity and thereby the BOLD signal in other, locally distant regions (e.g. insula and ACC). Future research should therefore complement pharmacological fMRI studies with other *in vivo* techniques that specifically monitor local changes in DA activity, such as molecular imaging with PET and SPECT (single photon emission computed tomography) in humans (see Cropley, Fujita, Innis, & Nathan, 2006), or fast-scan cyclic voltammetry and microdialysis in animals (Kehr & Yoshitake, 2013; Phillips, Robinson, Stuber, Carelli, & Wightman, 2002; Robinson, 2003). Another limitation relates to the point that the fMRI data acquired in this study were only analyzed using a standard mass-univariate approach, whereby alternative and more sophisticated methods are available that should be considered for further analysis. For example, multivariate approaches like multi-voxel pattern analysis (MVPA) might be more sensitive to detect DA drug effects on explore/exploit-related brain activity patterns than the applied voxel-wise approach (see Mahmoudi, Takerkart, Regragui, Boussaoud, & Brovelli, 2012; Norman, Polyn, Detre, & Haxby, 2006). Furthermore, as behavioral switching between exploitation and exploration is considered to involve the interplay of several brain regions (see 1.1.4), a functional connectivity analysis of the fMRI data based on dynamic causal modeling (DCM; Friston, Harrison, & Penny, 2003; Stephan & Friston, 2010) or psychophysiological interactions (PPI; Friston et al., 1997; O'Reilly et al., 2012) may provide more insight into the neural networks underlying explore/exploit behavior and their DA modulation.

Another limiting factor of this study is that the reported DA drug effects on insula and dACC activity resulted from an exploratory analysis of the fMRI data, which was performed only after the planned fMRI analysis failed to yield any significant drug effects. However, such post hoc exploratory analyses must be treated with caution, as they increase the likelihood to observe and report false positive findings – a problem that is often referred to as “data-dependent analysis” (Gelman & Lokenz, 2013) or “researchers degrees of freedom” (Simmons, Nelson, & Simonsohn, 2011). This problem is especially critical in brain imaging studies, where a large number of exploratory tests can be performed on the vast amount of data collected (Szucs & Ioannidis, 2017). Hence, the results of the exploratory fMRI analysis should only be regarded as preliminary evidence until replicated in further studies. Moreover, this study only investigated male participants, and future studies need to show whether these effects also generalize to the female population. In addition, it should be kept in mind that this work only focused on the role of DA in the explore/exploit trade-off, but did not consider critical interactive effects between DA and other neurotransmitters that are likely to contribute to this trade-off, especially NE and ACh (see 1.1.5). For instance, it has been suggested that the trade-off between exploitation and random exploration, which was not found to be affected by the DA drugs in this study, may be specifically regulated by the LC-NE system (see Aston-Jones & Cohen, 2005; Cohen et al., 2007; Doya, 2002; Ishii et al., 2002). Thus, a challenging prospect for future research will be to examine DA effects on explore/exploit behavior in a broader context, which also accounts for functional interactions between different neuromodulatory systems.

Another interesting prospect for future research arises from the finding that activity in the insula and dACC was increased during exploratory decisions and correlated with the overall reward uncertainty. Both these regions have been widely implicated in uncertainty coding, emotional processing, and mediating the effects of emotional arousal on (risky) decision making (see 6.3.3; e.g. Bush et al., 2000; Craig, 2002, 2009; Critchley, 2005; Dreher, 2013; Fukunaga et al., 2018; Preuschoff et al., 2008; Singer et al., 2009; Xue et al., 2010). Hence, the involvement of these regions in uncertainty coding and exploration might point to an interesting interpretation, according to which exploratory decisions may also be driven by emotional responses to accumulating uncertainty. In the literature, uncertainty-driven exploration is mostly considered to be rational and strategical, i.e. to account for the value of information in order to maximize rewards in the long term (see 1.1.3; e.g. Cogliati Dezza et al., 2017; Frank et al., 2009; Gershman, 2018; Wilson et al., 2014). This form of strategical exploration is considered to rely on a frontoparietal control network, in which the FPC tracks relevant decision parameters like the relative reward and uncertainty of alternative choice options in order to initiate exploratory actions whenever the available evidence supports such a decision (see 1.1.4; e.g. Badre et al., 2012; Boorman et al., 2009, 2011; Cavanagh et al., 2012). However, it might be assumed that also emotional arousal in response to increasing uncertainty or risk could play an important role in controlling exploratory behavior, and that these effects might be mediated by the insula and ACC. While the current study does not allow to directly test this assumption, future study could employ additional techniques to determine the extent to which exploratory actions are driven strategically vs. emotionally. For example, these studies could assess emotional responses to increasing uncertainty, e.g. by self-report or by recording physiological parameters like skin conductance and heart rate, and could test to what extent these responses correlate with activity in the insula/ACC and with exploratory behavior. Moreover, these studies may also provide more insight into the kinds of emotions triggered by growing uncertainty – e.g. the urge for taking a risk (Xue et al., 2010) or uncertainty-related anxiety (Hartley & Phelps, 2012; Hirsh, Mar, & Peterson, 2012) – and how these emotional responses may relate to individual differences in explore/exploit behavior and psychological constructs like risk/ambiguity aversion and novelty seeking (see Christopoulos et al., 2009; Costa et al., 2014; Hartley & Phelps, 2012; Levy, Snell, Nelson, Rustichini, & Glimcher, 2010; Payzan-LeNestour & Bossaerts, 2012). Additionally, such experiments may also use pupillometry to track cognitive processes during explore/exploit behavior, as pupil responses were recently found to indicate trial-by-trial changes in relevant decision parameters such as the expected value and uncertainty of reward (van Slooten, Jahfari, Knapen, & Theeuwes, 2018).

Finally, further research will be needed to better understand how alterations in the DA system contribute to dysfunctional decision making in psychiatric and neurological disorders. Recent studies have already started to investigate how explore/exploit behavior is altered in different clinical conditions, including depression (Blanco, Otto, Maddox, Beevers, & Love, 2013; Byrne, Norris, & Worthy, 2016), addiction (Addicott et al., 2013, 2014, 2015; Harlé et al., 2015; Morris et al., 2016),

schizophrenia (Strauss et al., 2011), and PD (Moustafa et al., 2008; Rutledge et al., 2009), as well as in healthy subjects in response to stress (Lenow et al., 2017), sleep deprivation (Glass et al., 2011), and childhood adversity (Humphreys et al., 2015). However, only few of these studies have specifically investigated the role of DA in altered explore/exploit behavior (e.g. Byrne et al., 2016; Moustafa et al., 2008; Rutledge et al., 2009), and more research is needed to bridge this gap. Eventually, the knowledge gained from these studies may also provide the basis for the development of new therapeutic interventions in the treatment of various disorders that involve a dysregulation of the explore/exploit trade-off.

6.7 Conclusion

The present study examined the causal role of DA in human explore/exploit behavior in a pharmacological fMRI approach, using the DA drugs L-dopa and haloperidol in a placebo-controlled, within-subjects design. First, the behavioral modeling results agree well with previous research, showing that humans use both random and directed exploration to solve the explore/exploit trade-off. More importantly, the findings of this study support the notion that DA is causally involved in this trade-off by regulating the extent to which subjects engage in uncertainty-driven exploration. Interestingly, the neuroimaging data of this study do not support the hypothesis that DA controls this trade-off by modulating the neural signatures of exploratory and exploitative decisions per se. In contrast, they provide preliminary evidence that DA may modulate uncertainty-related brain activity in a cortical control network comprising the insula and dACC, which may be involved in directing attention and behavior towards exploratory actions in the face of accumulating uncertainty. Future research should more closely examine the potential role of these regions in driving exploration based on emotional responses to increasing uncertainty, and further investigate how DA may be involved in this process.

7 Summary

A central aspect of many decision problems is the regulation of when to exploit, i.e. to choose a familiar option with a well-known reward, and when to explore, i.e. to choose an alternative option with an uncertain but potentially larger reward. This decision dilemma is commonly known as the explore/exploit trade-off, and a growing body of evidence suggests that dopamine (DA) may be tightly involved in regulating this trade-off. However, direct evidence for a causal role of DA in explore/exploit behavior is sparse and still lacking in humans. Therefore, the aim of this study was to directly test for DA effects on human explore/exploit behavior and its neural correlates in a pharmacological fMRI approach, using L-dopa (DA precursor) and haloperidol (DA antagonist) in a double-blind, placebo-controlled, within-subjects design. First, explore/exploit behavior, as assessed with the restless four-armed bandit task, was analyzed using different cognitive models of learning and decision making in a hierarchical Bayesian modeling approach. Among the tested models, choice behavior was best described by the Bayes-SM+EP model – a model that includes a Bayesian learning rule for tracking both the mean and variance (uncertainty) of the expected reward, combined with a modified softmax choice rule that captures both random and directed exploration along with choice perseveration. Using this model, it was found that directed (uncertainty-driven) exploration, as indexed by the φ parameter, was significantly reduced across subjects under L-dopa compared to placebo. In contrast, haloperidol did not significantly shift the φ parameter across subjects, but showed a tendency to reduce the group-level variance of this parameter relative to placebo and L-dopa. To examine drug effects on the neural level, choices were first classified as either exploitative (i.e. following the highest expected value) or exploratory, and the pattern of brain activity was compared between both types of choices. Across all drug conditions, exploratory choices were associated with higher activity in the frontopolar cortex (FPC) and intraparietal sulcus (IPS), consistent with previous studies which suggest that exploration is mediated via a frontoparietal control network. In contrast, exploitative choices showed higher activity in the orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC), brain areas that have previously been implicated in reward coding and exploitation. Surprisingly, no drug effects were found on these neural correlates of exploratory and exploitative choices, nor on striatal reward prediction error signaling. Yet, an exploratory analysis of the brain imaging data revealed that L-dopa reduced brain activity associated with the overall uncertainty in the insula and dorsal anterior cingulate cortex (dACC), areas that have been implicated in coding reward uncertainty and in mediating the effect of emotional arousal on risky decision making. Accordingly, by reducing uncertainty-related activity in these regions, L-dopa might have delayed the time point at which exploratory decisions are triggered in response to accumulating uncertainty. In conclusion, the results of this study support the notion that DA plays a causal role in human explore/exploit behavior. While more research is needed to reveal the underlying neural mechanisms involved in this process, first evidence suggests that DA may influence uncertainty-related activity in a cortical control network that guides attention and behavioral responses toward salient, uncertain choice options.

8 Zusammenfassung (German summary)

Zentraler Gegenstand vieler Entscheidungsprobleme ist die Frage, wann man „exploitet“, d.h. eine vertraute Option mit bekannter Belohnung wählt, und wann man „exploriert“, d.h. eine alternative Option mit einer ungewissen, aber möglicherweise höheren Belohnung wählt. Dieses Entscheidungsdilemma ist als „explore/exploit trade-off“ bekannt, und es gibt zunehmend Anhaltspunkte dafür, dass Dopamin (DA) eine zentrale Rolle in der Regulierung dieses trade-offs spielt. Direkte Evidenz dafür, dass DA kausal am explore/exploit Verhalten beteiligt ist, ist jedoch spärlich und fehlt bislang beim Menschen. Das Ziel dieser Studie war es daher, die kausale Rolle von DA in Bezug auf menschliches explore/exploit Verhalten und dessen neuronale Korrelate zu untersuchen. Hierzu wurde ein pharmakologischer fMRI Ansatz gewählt, bei dem L-Dopa (ein DA Vorläufer) und Haloperidol (ein DA Antagonist) in einem doppelt verblindeten, Placebo-kontrollierten Crossover-Studiendesign zur Anwendung kamen. Zunächst wurde explore/exploit Verhalten, das mittels der nicht-stationären vierarmigen Banditen-Aufgabe („restless four-armed bandit task“) untersucht wurde, unter Verwendung verschiedener kognitiver Modelle des Lernens und Entscheidens in einem hierarchischen Bayesianischen Modellierungsverfahren analysiert. Unter den getesteten Modellen wurde das Entscheidungsverhalten am besten durch das Bayes-SM+EP Model beschrieben. Dieses Modell kombiniert eine Bayesianische Lernregel, die sowohl den Mittelwert als auch die Varianz (Unsicherheit) der erwarteten Belohnung kontinuierlich aktualisiert, mit einer modifizierten Softmax-Entscheidungsregel, die sowohl zufällige als auch gerichtete Exploration und Perseveration erfasst. Anhand dieses Modells wurde festgestellt, dass gerichtete (unsicherheits-getriebene) Exploration, erfasst durch den φ Parameter, über alle Testpersonen hinweg unter L-Dopa gegenüber Placebo signifikant reduziert war. Dagegen zeigte sich unter Haloperidol keine signifikante Verschiebung des φ Parameters über alle Testpersonen hinweg, jedoch eine Tendenz zur Reduzierung der Gruppenvarianz dieses Parameters gegenüber Placebo und L-Dopa. Um die Effekte der dopaminergen Wirkstoffe auf der neuronalen Ebene zu untersuchen, wurden Entscheidungen zunächst als exploitativ (d.h. dem höchsten erwarteten Belohnungswert folgend) oder explorativ klassifiziert und das Muster der Gehirnaktivierung zwischen beiden Arten von Entscheidungen verglichen. Über alle experimentellen Bedingungen hinweg waren explorative Entscheidungen mit einer höheren Aktivität im frontopolaren Kortex (FPC) und intraparietalen Sulcus (IPS) assoziiert. Dieser Befund steht im Einklang mit früheren Studien, welche nahelegen, dass exploratives Verhalten über ein frontoparietales Kontrollnetzwerk vermittelt wird. Im Gegenzug dazu zeigten exploitative Entscheidungen eine höhere Aktivität im orbitofrontalen Kortex (OFC) und ventromedialen Präfrontalkortex (vmPFC) – Gehirnregionen, die zuvor mit der Kodierung von Belohnungen und exploitativem Verhalten in Verbindung gebracht wurden. Wider Erwarten zeigte sich kein signifikanter Einfluss der dopaminergen Substanzen auf die erwähnten neuronalen Korrelate explorativer und exploitativer Entscheidungen, noch auf die striatalen Korrelate des Belohnungs-Vorhersage-Fehlers. Hingegen ergab eine exploratorische Analyse der fMRI-Daten, dass L-Dopa zu einer Reduktion der mit der Gesamtunsicherheit (d.h. der Belohnungs-

unsicherheit über alle Entscheidungsoptionen hinweg) assoziierten Gehirnaktivität in der Insula und dem dorsalen anterioren cingulären Kortex (dACC) führt – Gehirnareale, die zuvor mit der Kodierung von Unsicherheit sowie mit der Vermittlung emotionaler Einflüsse auf riskantes Entscheidungsverhalten in Verbindung gebracht wurden. Demnach könnte vermutet werden, dass L-Dopa durch die Verringerung der unsicherheitsbezogenen Gehirnaktivität in diesen Arealen den Zeitpunkt verzögert hat, zu dem exploratorische Entscheidungen als Reaktion auf zunehmende Unsicherheit initiiert werden. Zusammenfassend stützen die Ergebnisse dieser Studie weitestgehend die Vorstellung, dass DA eine kausale Rolle im menschlichen explore/exploit Verhalten einnimmt. Wenngleich weitere Forschung nötig ist, um die zugrundeliegenden neuronalen Mechanismen dieser Prozesse aufzudecken, legen erste Befunde nahe, dass DA unsicherheitsbezogene Aktivität in einem kortikalen Kontrollnetzwerk beeinflusst, welches Aufmerksamkeit und Verhalten in Richtung salienter, unsicherer Entscheidungsoptionen lenkt.

9 Abbreviations

ACC	anterior cingulate cortex
AI	anterior insula
ANOVA	analysis of variance
BOLD	blood oxygenation level dependent
COMT	catechol-O-methyltransferase
CV	coefficient of variation
CV_{trials}	cross-validation over trials
DA	dopamine
dACC	dorsal anterior cingulate cortex
DAT	dopamine (active) transporter
DMN	default mode network
fMRI	functional magnetic resonance imaging
FPC	frontopolar cortex
FWE	familywise error
GLM	general linear model
HDI	highest density interval
IPS	intraparietal sulcus
LC	locus coeruleus
L-dopa	levo-3,4-dihydroxyphenylalanine
LM	linear model
lOFC	lateral orbitofrontal cortex
LOO	leave-one-out
NE	norepinephrine
OFC	orbitofrontal cortex
PCC	posterior cingulate cortex
PD	Parkinson's disease
PET	positron emission tomography
PFC	prefrontal cortex
QM	quadratic model
sEBR	spontaneous eye blink rate
SM	softmax
SM+E	softmax with exploration bonus
SM+EB	softmax with exploration bonus and perseveration bonus
SNc	substantia nigra pars compacta
tDCS	transcranial direct current stimulation
TMS	transcranial magnetic stimulation
vmPFC	ventromedial prefrontal cortex
VTA	ventral tegmental area
WMC	working memory capacity

10 List of symbols

α	learning rate
β	softmax parameter (inverse temperature)
δ	reward prediction error
ϑ	decay center
κ	Kalman gain
λ	decay parameter
Λ^x	group-level standard deviation of parameter x
M^x	group-level mean of parameter x
μ	mean of expected reward
ρ	perseveration bonus parameter
σ^2	variance of expected reward
σ_d^2	diffusion variance
σ_o^2	observation variance
φ	exploration bonus parameter
P	choice probability
r	reward
v	expected reward value

11 List of figures

12 List of tables

13 References

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S., & Zigmond, M. J. (1989). Differential Effect of Stress on In Vivo Dopamine Release in Striatum, Nucleus Accumbens, and Medial Frontal Cortex. *Journal of Neurochemistry*, *52*, 1655–1658. https://doi.org/10.1111/j.1471-4159.1989.tb09224.x

Abi-Dargham, A. (2004). Do we still believe in the dopamine hypothesis? New data bring new evidence. *The International Journal of Neuropsychopharmacology*, *7 Suppl 1*, S1-5. https://doi.org/10.1017/S1461145704004110

Abi-Dargham, A., & Moore, H. (2003). Prefrontal DA transmission at D1 receptors and the pathology of schizophrenia. *The Neuroscientist*, *9*, 404–416. https://doi.org/10.1177/1073858403252674

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage*, *31*, 790–795. https://doi.org/10.1016/j.neuroimage.2006.01.001

Adamantidis, A. R., Tsai, H.-C., Boutrel, B., Zhang, F., Stuber, G. D., Budygin, E. A., . . . de Lecea, L. (2011). Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *The Journal of Neuroscience*, *31*, 10829–10835. https://doi.org/10.1523/JNEUROSCI.2246-11.2011

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*, 53–63. https://doi.org/10.1136/jnnp-2015-310737

Adamson, T. A. (1995). Changes in blink rates of Nigerian schizophrenics treated with chlorpromazine. *West African Journal of Medicine*, *14*, 194–197.

Addicott, M. A., Pearson, J. M., Froeliger, B., Platt, M. L., & McClernon, F. J. (2014). Smoking automaticity and tolerance moderate brain activation during explore-exploit behavior. *Psychiatry Research*, *224*, 254–261. https://doi.org/10.1016/j.pscychresns.2014.10.014

Addicott, M. A., Pearson, J. M., Kaiser, N., Platt, M. L., & McClernon, F. J. (2015). Suboptimal foraging behavior: a new perspective on gambling. *Behavioral Neuroscience*, *129*, 656–665. https://doi.org/10.1037/bne0000082

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. *Neuropsychopharmacology*, *42*, 1931–1939. https://doi.org/10.1038/npp.2017.108

Addicott, M. A., Pearson, J. M., Wilson, J., Platt, M. L., & McClernon, F. J. (2013). Smoking and the bandit: a preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and Clinical Psychopharmacology*, *21*, 66–73. https://doi.org/10.1037/a0030843

Agarwal, D., Chen, B.-C., & Elango, P. (2009). Explore/exploit Schemes for Web Content Optimization. In W. Wang (Ed.), *Ninth IEEE International Conference on Data Mining, 2009: ICDM '09 ; Miami Beach, Florida, USA, 6-9 Dec. 2009* (pp. 1–10). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICDM.2009.52

Agostino, R., Bologna, M., Dinapoli, L., Gregori, B., Fabbrini, G., Accornero, N., & Berardelli, A. (2008). Voluntary, spontaneous, and reflex blinking in Parkinson's disease. *Movement Disorders*, *23*, 669–675. https://doi.org/10.1002/mds.21887

Ahn, W.-Y., Krawitz, A., Kim, W., Busmeyer, J. R., & Brown, J. W. (2011). A Model-Based fMRI Analysis with Hierarchical Bayesian Parameter Estimation. *Journal of Neuroscience, Psychology, and Economics*, *4*, 95–110. https://doi.org/10.1037/a0020684

Akbari Chermahini, S., & Hommel, B. (2010). The (b)link between creativity and dopamine: spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition*, *115*, 458–465. https://doi.org/10.1016/j.cognition.2010.03.007

Akbari Chermahini, S., & Hommel, B. (2012). More creative through positive mood? Not everyone! *Frontiers in Human Neuroscience*, *6*, 319. https://doi.org/10.3389/fnhum.2012.00319

Akil, M., Kolachana, B. S., Rothmond, D. A., Hyde, T. M., Weinberger, D. R., & Kleinman, J. E. (2003). Catechol-O-Methyltransferase Genotype and Dopamine Regulation in the Human Brain. *The Journal of Neuroscience*, *23*, 2008–2013. https://doi.org/10.1523/JNEUROSCI.23-06-02008.2003

Aksoy, D., Ortak, H., Kurt, S., Cevik, E., & Cevik, B. (2014). Central corneal thickness and its relationship to Parkinson's disease severity. *Canadian Journal of Ophthalmology*, *49*, 152–156. https://doi.org/10.1016/j.jcjo.2013.12.010

Alawieh, A., Zaraket, F. A., Li, J.-L., Mondello, S., Nokkari, A., Razafsha, M., . . . Kobeissy, F. H. (2012). Systems biology, bioinformatics, and biomarkers in neuropsychiatry. *Frontiers in Neuroscience*, *6*, 187. https://doi.org/10.3389/fnins.2012.00187

Alers, S., Bloembergen, D., Hennes, D., de Jong, S., Kaisers, M., Lemmens, N., . . . Weiss, G. (2011). Bee-inspired foraging in an embodied swarm (Demonstration). In K. Turner & P. Yolum (Eds.), *Proceedings of the 10th International Conference on Autonomous Agents & Multiagent Systems* (pp. 1311–1312). New York, NY: ACM Assoc. for Computing Machinery.

Almey, A., Milner, T. A., & Brake, W. G. (2015). Estrogen receptors in the central nervous system and their implication for dopamine-dependent cognition in females. *Hormones and Behavior*, *74*, 125–138. https://doi.org/10.1016/j.yhbeh.2015.06.010

Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annual Review of Neuroscience*, *25*, 189–220. https://doi.org/10.1146/annurev.neuro.25.112701.142922

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. *Prentice Hall Information and System Sciences Series*. Englewood Cliffs, NJ: Prentice-Hall.

Anderson, B. (2014). *Computational neuroscience and cognitive modelling: A student's introduction to methods and procedures*. London: Sage.

Anderson, E., & Nutt, J. (2011). The long-duration response to levodopa: phenomenology, potential mechanisms and clinical implications. *Parkinsonism & Related Disorders*, *17*, 587–592. https://doi.org/10.1016/j.parkreldis.2011.03.014

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*, 29–52. https://doi.org/10.1111/nyas.12360

Antinori, S., Fattore, L., Saba, P., Fratta, W., Gessa, G. L., & Devoto, P. (2018). Levodopa prevents the reinstatement of cocaine self-administration in rats via potentiation of dopamine release in the medial prefrontal cortex. *Addiction Biology*, *23*, 556–568. https://doi.org/10.1111/adb.12509

Arias-Carrión, O., & Pöppel, E. (2007). Dopamine, learning, and reward-seeking behavior. *Acta Neurobiologiae Experimentalis*, *67*, 481–488.

Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., & Pöppel, E. (2010). Dopaminergic reward system: a short integrative review. *International Archives of Medicine*, *3*, 24. https://doi.org/10.1186/1755-7682-3-24

Arnsten, A. F., Cai, J. X., Steere, J. C., & Goldman-Rakic, P. S. (1995). Dopamine D2 receptor mechanisms contribute to age-related cognitive decline: the effects of quinpirole on memory and motor performance in monkeys. *The Journal of Neuroscience*, *15*, 3429–3439.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*, 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Assadi, S. M., Yücel, M., & Pantelis, C. (2009). Dopamine modulates neural networks involved in effort-based decision-making. *Neuroscience and Biobehavioral Reviews*, *33*, 383–393. https://doi.org/10.1016/j.neubiorev.2008.10.010

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Audibert, J.-Y., Munos, R., & Szepesvári, C. (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, *410*, 1876–1902. https://doi.org/10.1016/j.tcs.2009.01.016

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, *47*, 235–256. https://doi.org/10.1023/A:1013689704352

Averbeck, B. B. (2015). Theory of choice in bandit, information sampling and foraging tasks. *PLoS Computational Biology*, *11*, e1004164. https://doi.org/10.1371/journal.pcbi.1004164

Ayano, G. (2016). Dopamine: Receptors, Functions, Synthesis, Pathways, Locations and Mental Disorders: Review of Literatures. *Journal of Mental Disorders and Treatment*, *2.* https://doi.org/10.4172/2471-271X.1000120

Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews. Neuroscience*, *13*, 572–586. https://doi.org/10.1038/nrn3289

Badgaiyan, R. D. (2011). Neurotransmitter Imaging: Basic Concepts and Future Perspectives. *Current Medical Imaging Reviews*, *7*, 98–103. https://doi.org/10.2174/157340511795445720

Badgaiyan, R. D. (2014). Imaging dopamine neurotransmission in live human brain. *Progress in Brain Research*, *211*, 165–182. https://doi.org/10.1016/B978-0-444-63425-2.00007-6

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*, 595–607. https://doi.org/10.1016/j.neuron.2011.12.025

Bai, J., Blot, K., Tzavara, E., Nosten-Bertrand, M., Giros, B., & Otani, S. (2014). Inhibition of dopamine transporter activity impairs synaptic depression in rat prefrontal cortex through over-stimulation of D1 receptors. *Cerebral Cortex*, *24*, 945–955. https://doi.org/10.1093/cercor/bhs376

Baldassano, C., & Leonard, N. E. (2009). Explore vs. Exploit: Task Allocation for Multi-robot Foraging. Retrieved from www.princeton.edu/~naomi/publications/2009/BalLeo09.pdf

Bannon, M. J., Wolf, M. E., & Roth, R. H. (1983). Pharmacology of dopamine neurons innervating the prefrontal, cingulate and piriform cortices. *European Journal of Pharmacology*, *92*, 119–125.

Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., & Rinaldi, F. (2000). Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research*, *93*, 145–151. https://doi.org/10.1016/S0165-1781(00)00108-6

Barkley-Levenson, E., & Galván, A. (2017). Eye blink rate predicts reward decisions in adolescents. *Developmental Science*, *20.* https://doi.org/10.1111/desc.12412

Barnett, J. H., Scoriels, L., & Munafò, M. R. (2008). Meta-analysis of the cognitive effects of the catechol-O-methyltransferase gene Val158/108Met polymorphism. *Biological Psychiatry*, *64*, 137–144. https://doi.org/10.1016/j.biopsych.2008.01.005

Barone, P. (2010). Neurotransmission in Parkinson's disease: beyond dopamine. *European Journal of Neurology*, *17*, 364–376. https://doi.org/10.1111/j.1468-1331.2009.02900.x

Bartkó, G., Herczeg, I., & Zádor, G. (1990). Blink rate response to haloperidol as possible predictor of therapeutic outcome. *Biological Psychiatry*, *27*, 113–115. https://doi.org/10.1016/0006-3223(90)90028-Z

Bartlett, E. J., Brodie, J. D., Simkowitz, P., Dewey, S. L., Rusinek, H., Wolf, A. P., . . . Wolkin, A. (1994). Effects of haloperidol challenge on regional cerebral glucose utilization in normal human subjects. *The American Journal of Psychiatry*, *151*, 681–686. https://doi.org/10.1176/ajp.151.5.681

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Baruzzi, A., Contin, M., Riva, R., Procaccianti, G., Albani, F., Tonello, C., . . . Martinelli, P. (1987). Influence of meal ingestion time on pharmacokinetics of orally administered levodopa in parkinsonian patients. *Clinical Neuropharmacology*, *10*, 527–537.

Beaulieu, J.-M., & Gainetdinov, R. R. (2011). The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacological Reviews*, *63*, 182–217. https://doi.org/10.1124/pr.110.002642

Beeler, J. A. (2012). Thorndike's Law 2.0: Dopamine and the Regulation of Thrift. *Frontiers in Neuroscience*, *6*, 116. https://doi.org/10.3389/fnins.2012.00116

Beeler, J. A., Cools, R., Luciana, M., Ostlund, S. B., & Petzinger, G. (2014). A kinder, gentler dopamine… highlighting dopamine's role in behavioral flexibility. *Frontiers in Neuroscience*, *8*, 4. https://doi.org/10.3389/fnins.2014.00004

Beeler, J. A., Daw, N., Frazier, C. R. M., & Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in Behavioral Neuroscience*, *4*, 170. https://doi.org/10.3389/fnbeh.2010.00170

Beeler, J. A., Frazier, C. R. M., & Zhuang, X. (2012). Putting desire on a budget: dopamine and energy expenditure, reconciling reward and resources. *Frontiers in Integrative Neuroscience*, *6*, 49. https://doi.org/10.3389/fnint.2012.00049

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221. https://doi.org/10.1038/nn1954

Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor. *Neuropsychopharmacology*, *38*, 1495–1503. https://doi.org/10.1038/npp.2013.48

Benabid, A. L., Chabardes, S., Mitrofanis, J., & Pollak, P. (2009). Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. *The Lancet Neurology*, *8*, 67–81. https://doi.org/10.1016/S1474-4422(08)70291-6

Benarroch, E. E. (2009). The locus ceruleus norepinephrine system: functional organization and potential clinical significance. *Neurology*, *73*, 1699–1704. https://doi.org/10.1212/WNL.0b013e3181c2937c

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: methodological and theoretical considerations. *Brain and Cognition*, *38*, 317–338. https://doi.org/10.1006/brcg.1998.1039

Berger, B., Gaspar, P., & Verney, C. (1991). Dopaminergic innervation of the cerebral cortex: unexpected differences between rodents and primates. *Trends in Neurosciences*, *14*, 21–27. https://doi.org/10.1016/0166-2236(91)90179-X

Berger-Tal, O., Nathan, J., Meron, E., & Saltz, D. (2014). The exploration-exploitation dilemma: a multidisciplinary framework. *PloS One*, *9*, e95693. https://doi.org/10.1371/journal.pone.0095693

Bergman, H., Feingold, A., Nini, A., Raz, A., Slovin, H., Abeles, M., & Vaadia, E. (1998). Physiological aspects of information processing in the basal ganglia of normal and parkinsonian primates. *Trends in Neurosciences*, *21*, 32–38. https://doi.org/10.1016/S0166-2236(97)01151-X

Bergman, H., Wichmann, T., Karmon, B., & DeLong, M. R. (1994). The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of parkinsonism. *Journal of Neurophysiology*, *72*, 507–520. https://doi.org/10.1152/jn.1994.72.2.507

Berman, B. D. (2011). Neuroleptic Malignant Syndrome: A Review for Neurohospitalists. *The Neurohospitalist*, *1*, 41–47. https://doi.org/10.1177/1941875210386491

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., & Seitelberger, F. (1973). Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations. *Journal of the Neurological Sciences*, *20*, 415–455.

Berridge, K. C. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. *The European Journal of Neuroscience*, *35*, 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, *86*, 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Berry, D. A., & Fristedt, B. (1985). *Bandit problems: Sequential Allocation of Experiments*. *Monographs on Statistics and Applied Probability*. Dordrecht: Springer. Retrieved from http://dx.doi.org/10.1007/978-94-015-3711-7

Bertler, A., Falck, B., Owman, C., & Rosengrenn, E. (1966). The localization of monoaminergic blood-brain barrier mechanisms. *Pharmacological Reviews*, *18*, 369–385.

Bhidayasiri, R., & Tarsy, D. (2012). Parkinson's Disease: "On-Off" Phenomenon. In *Movement Disorders: A Video Atlas: A Video Atlas* (pp. 14–15). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-60327-426-5_7

Bhidayasiri, R., & Truong, D. D. (2008). Motor complications in Parkinson disease: clinical manifestations and management. *Journal of the Neurological Sciences*, *266*, 204–215. https://doi.org/10.1016/j.jns.2007.08.028

Bilder, R. M., Volavka, J., Lachman, H. M., & Grace, A. A. (2004). The catechol-O-methyltransferase polymorphism: relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. *Neuropsychopharmacology*, *29*, 1943–1961. https://doi.org/10.1038/sj.npp.1300542

Bissonette, G. B., & Roesch, M. R. (2015). Development and function of the midbrain dopamine system: what we know and what we need to. *Genes, Brain, and Behavior*, *15*, 62–73. https://doi.org/10.1111/gbb.12257

Björklund, A., & Dunnett, S. B. (2007). Dopamine neuron systems in the brain: an update. *Trends in Neurosciences*, *30*, 194–202. https://doi.org/10.1016/j.tins.2007.03.006

Black, K. J., Piccirillo, M. L., Koller, J. M., Hseih, T., Wang, L., & Mintun, M. A. (2015). Levodopa effects on (11)Craclopride binding in the resting human brain. *F1000Research*, *4*, 23. https://doi.org/10.12688/f1000research.5672.1

Blackstone, C. (2009). Infantile parkinsonism-dystonia: a dopamine "transportopathy". *Journal of Clinical Investigation*, *10*, 167. https://doi.org/10.1172/JCI39632

Blanchard, T. C., & Gershman, S. J. (2018). Pure correlates of exploration and exploitation in the human brain. *Cognitive, Affective & Behavioral Neuroscience*, *18*, 117–126. https://doi.org/10.3758/s13415-017-0556-2

Blanco, N. J., Love, B. C., Cooper, J. A., McGeary, J. E., Knopik, V. S., & Maddox, W. T. (2015). A frontal dopamine system for reflective exploratory behavior. *Neurobiology of Learning and Memory*, *123*, 84–91. https://doi.org/10.1016/j.nlm.2015.05.004

Blanco, N. J., Otto, A. R., Maddox, W. T., Beevers, C. G., & Love, B. C. (2013). The influence of depression symptoms on exploratory decision-making. *Cognition*, *129*, 563–568. https://doi.org/10.1016/j.cognition.2013.08.018

Blin, O., Masson, G., Azulay, J. P., Fondarai, J., & Serratrice, G. (1990). Apomorphine-induced blinking and yawning in healthy volunteers. *British Journal of Clinical Pharmacology*, *30*, 769–773. https://doi.org/10.1111/j.1365-2125.1990.tb03848.x

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*, 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. *Brain Research. Cognitive Brain Research*, *25*, 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011

Bologna, M., Fasano, A., Modugno, N., Fabbrini, G., & Berardelli, A. (2012). Effects of subthalamic nucleus deep brain stimulation and L-DOPA on blinking in Parkinson's disease. *Experimental Neurology*, *235*, 265–272. https://doi.org/10.1016/j.expneurol.2012.02.004

Bond, A., & Lader, M. (1974). The use of analogue scales in rating subjective feelings. *British Journal of Medical Psychology*, *47*, 211–218. https://doi.org/10.1111/j.2044-8341.1974.tb02285.x

Bonnet, A.-M. (2000). Involvement of Non-Dopaminergic Pathways in Parkinson's Disease. *CNS Drugs*, *13*, 351–364. https://doi.org/10.2165/00023210-200013050-00005

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biology*, *9*, e1001093. https://doi.org/10.1371/journal.pbio.1001093

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*, 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958. https://doi.org/10.1038/ncomms15958

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*, 997–1003. https://doi.org/10.1038/nn.4573

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*, 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective & Behavioral Neuroscience*, *9*, 16–27. https://doi.org/10.3758/CABN.9.1.16

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Braver, T. S., & Bongiolatti, S. R. (2002). The role of frontopolar cortex in subgoal processing during working memory. *NeuroImage*, *15*, 523–536. https://doi.org/10.1006/nimg.2001.1019

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology*, *97*, 3036–3045. https://doi.org/10.1152/jn.01211.2006

Breitenstein, C., Korsukewitz, C., Flöel, A., Kretzschmar, T., Diederich, K., & Knecht, S. (2006). Tonic dopaminergic stimulation impairs associative learning in healthy subjects. *Neuropsychopharmacology*, *31*, 2552–2564. https://doi.org/10.1038/sj.npp.1301167

Brenner, G. M., & Stevens, C. W. (2013). *Pharmacology* (4th ed.). *Student consult*. Philadelphia, PA: Saunders/Elsevier.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, *14*, 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Brett, M., Penny, W., & Kiebel, S. (2006). Parametric procedures. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping.* San Diego: Academic Press.

Brezzi, M., & Lai, T. L. (2002). Optimal learning and experimentation in bandit problems. *Journal of Economic Dynamics and Control*, *27*, 87–108. https://doi.org/10.1016/S0165-1889(01)00028-8

Brisch, R., Saniotis, A., Wolf, R., Bielau, H., Bernstein, H.-G., Steiner, J., . . . Gos, T. (2014). The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Frontiers in Psychiatry*, *5*, 47. https://doi.org/10.3389/fpsyt.2014.00047

Brittain, J.-S., & Brown, P. (2014). Oscillations and the basal ganglia: motor control and beyond. *NeuroImage*, *85 Pt 2*, 637–647. https://doi.org/10.1016/j.neuroimage.2013.05.084

Brocka, M., Helbing, C., Vincenz, D., Scherf, T., Montag, D., Goldschmidt, J., . . . Lippert, M. (2018). Contributions of dopaminergic and non-dopaminergic neurons to VTA-stimulation induced neurovascular responses in brain reward circuits. *NeuroImage*, *177*, 88–97. https://doi.org/10.1016/j.neuroimage.2018.04.059

Brody, A. L., Mandelkern, M. A., London, E. D., Childress, A. R., Lee, G. S., Bota, R. G., . . . Jarvik, M. E. (2002). Brain metabolic changes during cigarette craving. *Archives of General Psychiatry*, *59*, 1162–1172.

Brody, A. L., Mandelkern, M. A., Olmstead, R. E., Scheibal, D., Hahn, E., Shiraga, S., . . . McCracken, J. T. (2006). Gene variants of brain dopamine pathways and smoking-induced dopamine release in the ventral caudate/nucleus accumbens. *Archives of General Psychiatry*, *63*, 808–816. https://doi.org/10.1001/archpsyc.63.7.808

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*, 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Brough, A., Isaac, M., & Chernev, A. (2008). The "Sticky Choice" Bias in Sequential Decision-Making. In *Advances in consumer research* (897-897). Duluth, Minn.: Assocation for Consumer Research.

Brown, P., Oliviero, A., Mazzone, P., Insola, A., Tonali, P., & Di Lazzaro, V. (2001). Dopamine Dependency of Oscillations between Subthalamic Nucleus and Pallidum in Parkinson's Disease. *The Journal of Neuroscience*, *21*, 1033–1038. https://doi.org/10.1523/JNEUROSCI.21-03-01033.2001

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., . . . Zald, D. H. (2010). Dopaminergic network differences in human impulsivity. *Science*, *329*, 532. https://doi.org/10.1126/science.1185778

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38. https://doi.org/10.1196/annals.1440.011

Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia*, *44*, 2594–2606. https://doi.org/10.1016/j.neuropsychologia.2005.10.011

Burke, C. J., Soutschek, A., Weber, S., Raja Beharelle, A., Fehr, E., Haker, H., & Tobler, P. N. (2018). Dopamine Receptor-Specific Contributions to the Computation of Value. *Neuropsychopharmacology*, *43*, 1415–1424. https://doi.org/10.1038/npp.2017.302

Burkey, A. R., Carstens, E., & Jasmin, L. (1999). Dopamine Reuptake Inhibition in the Rostral Agranular Insular Cortex Produces Antinociception. *The Journal of Neuroscience*, *19*, 4169–4179. https://doi.org/10.1523/JNEUROSCI.19-10-04169.1999

Burkhardt, J. M., Constantinidis, C., Anstrom, K. K., Roberts, D. C. S., & Woodward, D. J. (2007). Synchronous oscillations and phase reorganization in the basal ganglia during akinesia induced by high-dose haloperidol. *The European Journal of Neuroscience*, *26*, 1912–1924. https://doi.org/10.1111/j.1460-9568.2007.05813.x

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*, 215–222.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, *14*, 365–376. https://doi.org/10.1038/nrn3475

Bymaster, F., Perry, K. W., Nelson, D. L., Wong, D. T., Rasmussen, K., Moore, N. A., & Calligaro, D. O. (1999). Olanzapine: a basic science update. *British Journal of Psychiatry*, *174*, 36–40. https://doi.org/10.1192/S0007125000293653

Byrne, K. A., Norris, D. D., & Worthy, D. A. (2016). Dopamine, depressive symptoms, and decision-making: the relationship between spontaneous eye blink rate and depressive symptoms predicts Iowa Gambling Task performance. *Cognitive, Affective & Behavioral Neuroscience*, *16*, 23–36. https://doi.org/10.3758/s13415-015-0377-0

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., . . . Greengard, P. (2000). Dopamine and cAMP-Regulated Phosphoprotein 32 kDa Controls Both Striatal Long-Term Depression and Long-Term Potentiation, Opposing Forms of Synaptic Plasticity. *The Journal of Neuroscience*, *20*, 8443–8451. https://doi.org/10.1523/JNEUROSCI.20-22-08443.2000

Caldú, X., Vendrell, P., Bartrés-Faz, D., Clemente, I., Bargalló, N., Jurado, M. A., . . . Junqué, C. (2007). Impact of the COMT Val108/158 Met and DAT genotypes on prefrontal function in healthy subjects. *NeuroImage*, *37*, 1437–1444. https://doi.org/10.1016/j.neuroimage.2007.06.021

Calipari, E. S., & Ferris, M. J. (2013). Amphetamine mechanisms and actions at the dopamine terminal revisited. *The Journal of Neuroscience*, *33*, 8923–8925. https://doi.org/10.1523/JNEUROSCI.1033-13.2013

Camps, M., Cortés, R., Gueye, B., Probst, A., & Palacios, J. M. (1989). Dopamine receptors in human brain: Autoradiographic distribution of D2 sites. *Neuroscience*, *28*, 275–290. https://doi.org/10.1016/0306-4522(89)90179-6

Carboni, E., Silvagni, A., Vacca, C., & Di Chiara, G. (2006). Cumulative effect of norepinephrine and dopamine carrier blockade on extracellular dopamine increase in the nucleus accumbens shell, bed nucleus of stria terminalis and prefrontal cortex. *Journal of Neurochemistry*, *96*, 473–481. https://doi.org/10.1111/j.1471-4159.2005.03556.x

Carey, R. J., Dai, H., Huston, J. P., Pinheiro-Carrera, M., Schwarting, R.K.W., & Tomaz, C. (1995). L-DOPA metabolism in cortical and striatal tissues in an animal model of Parkinsonism. *Brain Research Bulletin*, *37*, 295–299. https://doi.org/10.1016/0361-9230(95)00019-B

Carey, R. J., Pinheiro-Carrera, M., Dai, H., Tomaz, C., & Huston, J. P. (1995). L-DOPA and psychosis: Evidence for L-DOPA-induced increases in prefrontal cortex dopamine and in serum corticosterone. *Biological Psychiatry*, *38*, 669–676. https://doi.org/10.1016/0006-3223(94)00378-5

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76.* https://doi.org/10.18637/jss.v076.i01

Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, *7(3)*, 863–882. Retrieved from http://www.redalyc.org/articulo.oa?id=33770319

Carvalho, M. M., Campos, F. L., Marques, M., Soares-Cunha, C., Kokras, N., Dalla, C., . . . Salgado, A. J. (2017). Effect of Levodopa on Reward and Impulsivity in a Rat Model of Parkinson's Disease. *Frontiers in Behavioral Neuroscience*, *11*, 145. https://doi.org/10.3389/fnbeh.2017.00145

Cass, W. A., & Gerhardt, G. A. (1995). In vivo assessment of dopamine uptake in rat medial prefrontal cortex: comparison with dorsal striatum and nucleus accumbens. *Journal of Neurochemistry*, *65*, 201–207.

Cass, W. A., Zahniser, N. R., Flach, K. A., & Gerhardt, G. A. (1993). Clearance of Exogenous Dopamine in Rat Dorsal Striatum and Nucleus Accumbens: Role of Metabolism and Effects of Locally Applied Uptake Inhibitors. *Journal of Neurochemistry*, *61*, 2269–2278. https://doi.org/10.1111/j.1471-4159.1993.tb07469.x

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex*, *22*, 2575–2586. https://doi.org/10.1093/cercor/bhr332

Cavanagh, J. F., Masters, S. E., Bath, K., & Frank, M. J. (2014). Conflict acts as an implicit cost in reinforcement learning. *Nature Communications*, *5*, 5394. https://doi.org/10.1038/ncomms6394

Cesa-Bianchi, N., & Fischer, P. (1998). Finite-time Regret Bounds for the Multiarmed Bandit Problem. In *In 5th International Conference on Machine Learning* (pp. 100–108). Morgan Kaufmann.

Chang, C. Y., Gardner, M. P. H., Conroy, J. C., Whitaker, L. R., & Schoenbaum, G. (2018). Brief, But Not Prolonged, Pauses in the Firing of Midbrain Dopamine Neurons Are Sufficient to Produce a Conditioned Inhibitor. *The Journal of Neuroscience*, *38*, 8822–8830. https://doi.org/10.1523/JNEUROSCI.0144-18.2018

Chapuis, S., Ouchchane, L., Metz, O., Gerbaud, L., & Durif, F. (2005). Impact of the motor complications of Parkinson's disease on the quality of life. *Movement Disorders*, *20*, 224–230. https://doi.org/10.1002/mds.20279

Chavhan, G. B., Babyn, P. S., Thomas, B., Shroff, M. M., & Haacke, E. M. (2009). Principles, techniques, and applications of T2*-based MR imaging and its special applications. *Radiographics*, *29*, 1433–1449. https://doi.org/10.1148/rg.295095034

Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., & Cox, R. W. (2012). Fmri group analysis combining effect estimates and their variances. *NeuroImage*, *60*, 747–765. https://doi.org/10.1016/j.neuroimage.2011.12.060

Chen, J., Song, J., Yuan, P., Tian, Q., Ji, Y., Ren-Patterson, R., . . . Weinberger, D. R. (2011). Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: implications for drug development. *The Journal of Biological Chemistry*, *286*, 34752–34760. https://doi.org/10.1074/jbc.M111.262790

Chen, N. N., & Pan, W. H. (2000). Regulatory effects of D2 receptors in the ventral tegmental area on the mesocorticolimbic dopaminergic pathway. *Journal of Neurochemistry*, *74*, 2576–2582.

Chen, Y.-C. I., Choi, J.-K., Andersen, S. L., Rosen, B. R., & Jenkins, B. G. (2005). Mapping dopamine D2/D3 receptor function using pharmacological magnetic resonance imaging. *Psychopharmacology*, *180*, 705–715. https://doi.org/10.1007/s00213-004-2034-0

Chikama, M., McFarland, N. R., Amaral, D. G., & Haber, S. N. (1997). Insular Cortical Projections to Functional Regions of the Striatum Correlate with Cortical Cytoarchitectonic Organization in the Primate. *The Journal of Neuroscience*, *17*, 9686–9705. https://doi.org/10.1523/JNEUROSCI.17-24-09686.1997

Christoff, K., & Gabrieli, J.D.E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168–186. https://doi.org/10.3758/BF03331976

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., & Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *The Journal of Neuroscience*, *29*, 12574–12583. https://doi.org/10.1523/JNEUROSCI.2614-09.2009

Ciliax, B. J., Drash, G. W., Staley, J. K., Haber, S., Mobley, C. J., Miller, G. W., . . . Levey, A. I. (1999). Immunocytochemical localization of the dopamine transporter in human brain. *The Journal of Comparative Neurology*, *409*, 38–56. https://doi.org/10.1002/(SICI)1096-9861(19990621)409:1<38::AID-CNE4>3.0.CO;2-1

Clarke, H. F., Cardinal, R. N., Rygula, R., Hong, Y. T., Fryer, T. D., Sawiak, S. J., . . . Roberts, A. C. (2014). Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. *The Journal of Neuroscience*, *34*, 7663–7676. https://doi.org/10.1523/JNEUROSCI.0718-14.2014

Coffeen, U., López-Avila, A., Ortega-Legaspi, J. M., del Angel, R., López-Muñoz, F. J., & Pellicer, F. (2008). Dopamine receptors in the anterior insular cortex modulate long-term nociception in the rat. *European Journal of Pain*, *12*, 535–543. https://doi.org/10.1016/j.ejpain.2007.08.008

Coffeen, U., Ortega-Legaspi, J. M., de Gortari, P., Simón-Arceo, K., Jaimes, O., Amaya, M. I., & Pellicer, F. (2010). Inflammatory nociception diminishes dopamine release and increases dopamine D2 receptor mRNA in the rat's insular cortex. *Molecular Pain*, *6*, 75. https://doi.org/10.1186/1744-8069-6-75

Coffeen, U., Ortega-Legaspi, J. M., & Pellicer, F. (2012). Dopamine and pain modulation in the insular cortex. In E. Kudo & Y. Fujii (Eds.), *Neuroscience research progress. Dopamine: Functions, regulation and health effects* (pp. 235–247). New York: Nova Science Publishers.

Cogliati Dezza, I., Yu, A. J., Cleeremans, A., & Alexander, W. (2017). Learning the value of information and reward over time when solving exploration-exploitation problems. *Scientific Reports*, *7*, 16919. https://doi.org/10.1038/s41598-017-17237-w

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*, 933–942. https://doi.org/10.1098/rstb.2007.2098

Cohen, M. X., Krohn-Grimberghe, A., Elger, C. E., & Weber, B. (2007). Dopamine gene predicts the brain's response to dopaminergic drug. *The European Journal of Neuroscience*, *26*, 3652–3660. https://doi.org/10.1111/j.1460-9568.2007.05947.x

Constantino, S. M., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective & Behavioral Neuroscience*, *15*, 837–853. https://doi.org/10.3758/s13415-015-0350-y

Contin, M., & Martinelli, P. (2010). Pharmacokinetics of levodopa. *Journal of Neurology*, *257*, S253-61. https://doi.org/10.1007/s00415-010-5728-8

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. https://doi.org/10.3758/BF03196772

Cook, Z., Franks, D. W., & Robinson, E. J. H. (2013). Exploration versus exploitation in polydomous ant colonies. *Journal of Theoretical Biology*, *323*, 49–56. https://doi.org/10.1016/j.jtbi.2013.01.022

Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neuroscience and Biobehavioral Reviews*, *30*, 1–23. https://doi.org/10.1016/j.neubiorev.2005.03.024

Cools, R. (2011). Dopaminergic control of the striatum for high-level cognition. *Current Opinion in Neurobiology*, *21*, 402–407. https://doi.org/10.1016/j.conb.2011.04.002

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2003). L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. *Neuropsychologia*, *41*, 1431–1441. https://doi.org/10.1016/S0028-3932(03)00117-9

Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry*, *69*, e113-25. https://doi.org/10.1016/j.biopsych.2011.03.028

Cools, R., Frank, M. J., Gibbs, S. E. B., Miyakawa, A., Jagust, W., & D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *The Journal of Neuroscience*, *29*, 1538–1543. https://doi.org/10.1523/JNEUROSCI.4467-08.2009

Cools, R., Gibbs, S. E. B., Miyakawa, A., Jagust, W., & D'Esposito, M. (2008). Working memory capacity predicts dopamine synthesis capacity in the human striatum. *The Journal of Neuroscience*, *28*, 1208–1212. https://doi.org/10.1523/JNEUROSCI.4475-07.2008

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain (Ph.D. Thesis). McGill University, Montreal, Canada.

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2014). Dopamine modulates novelty seeking behavior during decision making. *Behavioral Neuroscience*, *128*, 556–566. https://doi.org/10.1037/a0037128

Cox, S. M. L., Frank, M. J., Larcher, K., Fellows, L. K., Clark, C. A., Leyton, M., & Dagher, A. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *NeuroImage*, *109*, 95–101. https://doi.org/10.1016/j.neuroimage.2014.12.070

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, *3*, 655–666. https://doi.org/10.1038/nrn894

Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, *10*, 59–70. https://doi.org/10.1038/nrn2555

Crevoisier, C., Hoevels, B., Zürcher, G., & Da Prada, M. (1987). Bioavailability of L-Dopa after Madopar HBS administration in healthy volunteers. *European Neurology*, *27 Suppl 1*, 36–46. https://doi.org/10.1159/000116173

Critchley, H. D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *The Journal of Comparative Neurology*, *493*, 154–166. https://doi.org/10.1002/cne.20749

Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, *77*, 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Critchley, H. D., Mathiast, C. J., & Dolan, R. J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. *NeuroImage*, *13*, 392. https://doi.org/10.1016/S1053-8119(01)91735-5

Cropley, V. L., Fujita, M., Innis, R. B., & Nathan, P. J. (2006). Molecular imaging of the dopaminergic system and its association with human cognitive function. *Biological Psychiatry*, *59*, 898–907. https://doi.org/10.1016/j.biopsych.2006.03.004

Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, *16*, 205–212. https://doi.org/10.1016/j.conb.2006.03.005

Currie, S., Hoggard, N., Craven, I. J., Hadjivassiliou, M., & Wilkinson, I. D. (2013). Understanding MRI: basic MR physics for physicians. *Postgraduate Medical Journal*, *89*, 209–223. https://doi.org/10.1136/postgradmedj-2012-131342

Dalley, J. W., Fryer, T. D., Brichard, L., Robinson, E. S. J., Theobald, D. E. H., Lääne, K., . . . Robbins, T. W. (2007). Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. *Science*, *315*, 1267–1270. https://doi.org/10.1126/science.1137073

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews. Neuroscience*, 143–152. https://doi.org/10.1038/nrn3403

D'Ambrosio, A., Zivkovic, B., & Bartholini, G. (1982). [3H]haloperidol labels brain dopamine receptors after its injection into the internal carotid artery of the rat. *Brain Research*, *238*, 470–474. https://doi.org/10.1016/0006-8993(82)90125-1

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Dang, J., Xiao, S., Liu, Y., Jiang, Y., & Mao, L. (2016). Individual differences in dopamine level modulate the ego depletion effect. *International Journal of Psychophysiology*, *99*, 121–124. https://doi.org/10.1016/j.ijpsycho.2015.11.013

Dang, L. C., Samanez-Larkin, G. R., Castrellon, J. J., Perkins, S. F., Cowan, R. L., Newhouse, P. A., & Zald, D. H. (2017). Spontaneous Eye Blink Rate (EBR) Is Uncorrelated with Dopamine D2 Receptor Availability and Unmodulated by Dopamine Agonism in Healthy Adults. *ENeuro*, *4.* https://doi.org/10.1523/ENEURO.0211-17.2017

Davis, J. M., Chen, N., & Glick, I. D. (2003). A meta-analysis of the efficacy of second-generation antipsychotics. *Archives of General Psychiatry*, *60*, 553–564. https://doi.org/10.1001/archpsyc.60.6.553

Davis, K. L., Kahn, R. S., Ko, G., & Davidson, M. (1991). Dopamine in schizophrenia: a review and reconceptualization. *The American Journal of Psychiatry*, *148*, 1474–1486. https://doi.org/10.1176/ajp.148.11.1474

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision Making, Affect, and Learning* (pp. 3–38). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Daw, N. D. (2014). Advanced Reinforcement Learning. In *Neuroeconomics* (pp. 299–320). Elsevier. https://doi.org/10.1016/B978-0-12-416008-8.00016-4

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*, 199–204. https://doi.org/10.1016/j.conb.2006.03.006

Daw, N. D., O'Doherty, J., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879. https://doi.org/10.1038/nature04766

Dayan, P., & Sejnowski, T. J. (1996). Exploration bonuses and dual control. *Machine Learning*, *25*, 5–22. https://doi.org/10.1007/BF00115298

De Mei, C., Ramos, M., Iitaka, C., & Borrelli, E. (2009). Getting specialized: presynaptic and postsynaptic dopamine D2 receptors. *Current Opinion in Pharmacology*, *9*, 53–58. https://doi.org/10.1016/j.coph.2008.12.002

De Vries, M. H., Ulte, C., Zwitserlood, P., Szymanski, B., & Knecht, S. (2010). Increasing dopamine levels in the brain improves feedback-based procedural learning in healthy participants: an artificial-grammar-learning experiment. *Neuropsychologia*, *48*, 3193–3197. https://doi.org/10.1016/j.neuropsychologia.2010.06.024

Derogatis, L. R. (1992). *The Symptom Checklist-90-revised*. Minneapolis, MN: NCS Assessments.

Deuschl, G., & Goddemeier, C. (1998). Spontaneous and reflex activity of facial muscles in dystonia, Parkinson's disease, and in normal subjects. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*, 320–324. https://doi.org/10.1136/jnnp.64.3.320

Devoto, P., Fattore, L., Antinori, S., Saba, P., Frau, R., Fratta, W., & Gessa, G. L. (2016). Elevated dopamine in the medial prefrontal cortex suppresses cocaine seeking via D1 receptor overstimulation. *Addiction Biology*, *21*, 61–71. https://doi.org/10.1111/adb.12178

Dewey, R. B. (2004). Management of motor complications in Parkinson's disease. *Neurology*, *62*, S3-S7. https://doi.org/10.1212/WNL.62.6_suppl_4.S3

Dodel, R. C., Berger, K., & Oertel, W. H. (2001). Health-related quality of life and healthcare utilisation in patients with Parkinson's disease: impact of motor fluctuations and dyskinesias. *PharmacoEconomics*, *19*, 1013–1038. https://doi.org/10.2165/00019053-200119100-00004

Dold, M., Samara, M. T., Li, C., Tardy, M., & Leucht, S. (2015). Haloperidol versus first-generation antipsychotics for the treatment of schizophrenia and other psychotic disorders. *The Cochrane Database of Systematic Reviews*, *1*, CD009831. https://doi.org/10.1002/14651858.CD009831.pub2

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106. https://doi.org/10.1016/j.cobeha.2014.10.007

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*, 365–378. https://doi.org/10.1016/j.neuron.2004.09.009

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*, 495–506. https://doi.org/10.1016/S0893-6080(02)00044-8

Dreher, J.-C. (2013). Neural coding of computational factors affecting decision making. *Progress in Brain Research*, *202*, 289–320. https://doi.org/10.1016/B978-0-444-62604-2.00016-2

Dreher, J.-C., Kohn, P., & Berman, K. F. (2006). Neural coding of distinct statistical properties of reward information in humans. *Cerebral Cortex*, *16*, 561–573. https://doi.org/10.1093/cercor/bhj004

Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K.-P., & Brocke, B. (2005). Dopamine and cognitive control: the influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility. *Behavioral Neuroscience*, *119*, 483–490. https://doi.org/10.1037/0735-7044.119.2.483

Druzin, M. Y., Kurzina, N. P., Malinina, E. P., & Kozlov, A. P. (2000). The effects of local application of D2 selective dopaminergic drugs into the medial prefrontal cortex of rats in a delayed spatial choice task. *Behavioural Brain Research*, *109*, 99–111. https://doi.org/10.1016/S0166-4328(99)00166-7

Durstewitz, D., & Seamans, J. K. (2002). The computational role of dopamine D1 receptors in working memory. *Neural Networks*, *15*, 561–572.

Duvarci, S., Simpson, E. H., Schneider, G., Kandel, E., Roeper, J., & Sigurdsson, T. (2018). Impaired recruitment of dopamine neurons during working memory in mice with striatal D2 receptor overexpression. *Nature Communications*, *9*, 2822. https://doi.org/10.1038/s41467-018-05214-4

Ebert, D., Albert, R., Hammon, G., Strasser, B., May, A., & Merz, A. (1996). Eye-blink rates and depression. Is the antidepressant effect of sleep deprivation mediated by the dopamine system? *Neuropsychopharmacology*, *15*, 332–339. https://doi.org/10.1016/0893-133X(95)00237-8

Efron, B., & Morris, C. (1977). Stein's Paradox in Statistics. *Scientific American*, *236*, 119–127. https://doi.org/10.1038/scientificamerican0577-119

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., . . . Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 6917–6922. https://doi.org/10.1073/pnas.111134598

Egerton, A., Mehta, M. A., Montgomery, A. J., Lappin, J. M., Howes, O. D., Reeves, S. J., . . . Grasby, P. M. (2009). The dopaminergic basis of human behaviors: A review of molecular imaging studies. *Neuroscience and Biobehavioral Reviews*, *33*, 1109–1132. https://doi.org/10.1016/j.neubiorev.2009.05.005

Eisenegger, C., Naef, M., Linssen, A., Clark, L., Gandamaneni, P. K., Müller, U., & Robbins, T. W. (2014). Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology*, *39*, 2366–2375. https://doi.org/10.1038/npp.2014.84

Elston, T. W., & Bilkey, D. K. (2017). Anterior Cingulate Cortex Modulation of the Ventral Tegmental Area in an Effort Task. *Cell Reports*, *19*, 2220–2230. https://doi.org/10.1016/j.celrep.2017.05.062

Evans, A. H., Pavese, N., Lawrence, A. D., Tai, Y. F., Appel, S., Doder, M., . . . Piccini, P. (2006). Compulsive drug use linked to sensitized ventral striatal dopamine transmission. *Annals of Neurology*, *59*, 852–858. https://doi.org/10.1002/ana.20822

Evans, L. J., & Raine, N. E. (2014). Foraging errors play a role in resource exploration by bumble bees (Bombus terrrestris). *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, *200*, 475–484. https://doi.org/10.1007/s00359-014-0905-3

Even-Dar, E., Mannor, S., & Mansour, Y. (2002). PAC Bounds for Multi-armed Bandit and Markov Decision Processes. In K. Kivinen & R. H. Sloan (Eds.), *Lecture Notes in Artificial Intelligence Lecture Notes in Computer Science: Vol. 2375. Computational Learning Theory: 15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 8-10, 2002: Proceedings* (Vol. 2375, pp. 255–270). New York: Springer. https://doi.org/10.1007/3-540-45435-7_18

Fahn, S. (2006). Levodopa in the treatment of Parkinson's disease. *Journal of Neural Transmission. Supplementum*, 1–15.

Fahn, S., Jankovic, J., & Hallett, M. (2011). Medical treatment of Parkinson disease. In *Principles and Practice of Movement Disorders* (pp. 119–156). Elsevier. https://doi.org/10.1016/B978-1-4377-2369-4.00006-8

Fallon, S. J., Smulders, K., Esselink, R. A., van de Warrenburg, B. P., Bloem, B. R., & Cools, R. (2015). Differential optimal dopamine levels for set-shifting and working memory in Parkinson's disease. *Neuropsychologia*, *77*, 42–51. https://doi.org/10.1016/j.neuropsychologia.2015.07.031

Fang, J. Y., & Tolleson, C. (2017). The role of deep brain stimulation in Parkinson's disease: an overview and update on new developments. *Neuropsychiatric Disease and Treatment*, *13*, 723–732. https://doi.org/10.2147/NDT.S113998

Farrell, S. M., Tunbridge, E. M., Braeutigam, S., & Harrison, P. J. (2012). COMT Val(158)Met genotype determines the direction of cognitive effects produced by catechol-O-methyltransferase inhibition. *Biological Psychiatry*, *71*, 538–544. https://doi.org/10.1016/j.biopsych.2011.12.023

Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Annals of the New York Academy of Sciences*, *1239*, 51–58. https://doi.org/10.1111/j.1749-6632.2011.06229.x

Fiorillo, C. D., Newsome, W. T., & Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. *Nature Neuroscience*, *11*, 966–973. https://doi.org/10.1038/nn.2159

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*, 1898–1902. https://doi.org/10.1126/science.1077349

Fitzgerald, T. H. B., Seymour, B., Bach, D. R., & Dolan, R. J. (2010). Differentiable neural substrates for learned and described value and risk. *Current Biology*, *20*, 1823–1829. https://doi.org/10.1016/j.cub.2010.08.048

Fleckenstein, A. E., Volz, T. J., Riddle, E. L., Gibb, J. W., & Hanson, G. R. (2007). New insights into the mechanism of action of amphetamines. *Annual Review of Pharmacology and Toxicology*, *47*, 681–698. https://doi.org/10.1146/annurev.pharmtox.47.120505.105140

Floel, A., Breitenstein, C., Hummel, F., Celnik, P., Gingert, C., Sawaki, L., . . . Cohen, L. G. (2005). Dopaminergic influences on formation of a motor memory. *Annals of Neurology*, *58*, 121–130. https://doi.org/10.1002/ana.20536

Floel, A., Garraux, G., Xu, B., Breitenstein, C., Knecht, S., Herscovitch, P., & Cohen, L. G. (2008). Levodopa increases memory encoding and dopamine release in the striatum in the elderly. *Neurobiology of Aging*, *29*, 267–279. https://doi.org/10.1016/j.neurobiolaging.2006.10.009

Floresco, S. B. (2013). Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. *Frontiers in Neuroscience*, *7*, 62. https://doi.org/10.3389/fnins.2013.00062

Floresco, S. B., & Magyar, O. (2006). Mesocortical dopamine modulation of executive functions: beyond working memory. *Psychopharmacology*, *188*, 567–585. https://doi.org/10.1007/s00213-006-0404-5

Floresco, S. B., Magyar, O., Ghods-Sharifi, S., Vexelman, C., & Tse, M. T. L. (2006). Multiple dopamine receptor subtypes in the medial prefrontal cortex of the rat regulate set-shifting. *Neuropsychopharmacology*, *31*, 297–309. https://doi.org/10.1038/sj.npp.1300825

Floresco, S. B., West, A. R., Ash, B., Moore, H., & Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature Neuroscience*, *6*, 968–973. https://doi.org/10.1038/nn1103

Ford, C. P. (2014). The role of D2-autoreceptors in regulating dopamine neuron activity and transmission. *Neuroscience*, *282*, 13–22. https://doi.org/10.1016/j.neuroscience.2014.01.025

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226–236. https://doi.org/10.3758/s13421-014-0461-7

Fountas, Z., & Shanahan, M. (2017). The role of cortical oscillations in a spiking neural network model of the basal ganglia. *PloS One*, *12*, e0189109. https://doi.org/10.1371/journal.pone.0189109

Fox, P. T., & Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences of the United States of America*, *83*, 1140–1144.

Fox, P. T., Raichle, M. E., Mintun, M. A., & Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, *241*, 462–464.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72. https://doi.org/10.1162/0898929052880093

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*, 1062–1068. https://doi.org/10.1038/nn.2342

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, *120*, 497–517. https://doi.org/10.1037/0735-7044.120.3.497

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*, 1940–1943. https://doi.org/10.1126/science.1102941

Franke, G. H. (1995). *Die Symptom-Checkliste von Derogatis (SCL-90-R) – Deutsche Version – Manual*. Göttingen, Germany: Beltz Test GmbH.

Franke, G. H. (2000). *Die Symptom-Checkliste von Derogatis (SCL-90-R) – Deutsche Version – Manual* (2nd ed.). Göttingen, Germany: Beltz Test GmbH.

Frazier, P. I., Powell, W. B., & Dayanik, S. (2008). A Knowledge-Gradient Policy for Sequential Information Collection. *SIAM Journal on Control and Optimization*, *47*, 2410–2439. https://doi.org/10.1137/070693424

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (Eds.). (2006). *Statistical Parametric Mapping*. San Diego: Academic Press.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*, 218–229. https://doi.org/10.1006/nimg.1997.0291

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2(4)*, 189–210. https://doi.org/10.1002/hbm.460020402

Friston, K. J., Stephan, K. E., Lund, T. E., Morcom, A., & Kiebel, S. (2005). Mixed-effects and fMRI studies. *NeuroImage*, *24*, 244–252. https://doi.org/10.1016/j.neuroimage.2004.08.055

Fukunaga, R., Purcell, J. R., & Brown, J. W. (2018). Discriminating Formal Representations of Risk in Anterior Cingulate Cortex and Inferior Frontal Gyrus. *Frontiers in Neuroscience*, *12*, 553. https://doi.org/10.3389/fnins.2018.00553

Fum, D., Missier, F. D., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, *8*, 135–142. https://doi.org/10.1016/j.cogsys.2007.07.001

Gabry, J., & Goodrich, B. (2018). Posterior uncertainty intervals. Retrieved from http://mc-stan.org/rstanarm/reference/posterior_interval.stanreg.html

Galvan, A., & Wichmann, T. (2008). Pathophysiology of Parkinsonism. *Clinical Neurophysiology*, *119*, 1459–1474. https://doi.org/10.1016/j.clinph.2008.03.017

Garavan, H., Pankiewicz, J., Bloom, A., Cho, J. K., Sperry, L., Ross, T. J., . . . Stein, E. A. (2000). Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli. *The American Journal of Psychiatry*, *157*, 1789–1798. https://doi.org/10.1176/appi.ajp.157.11.1789

Garris, P. A., Budygin, E.A., Phillips, P.E.M., Venton, B.J., Robinson, D.L., Bergstrom, B.P., . . . Wightman, R. M. (2003). A role for presynaptic mechanisms in the actions of nomifensine and haloperidol. *Neuroscience*, *118*, 819–829. https://doi.org/10.1016/S0306-4522(03)00005-8

Garris, P. A., & Wightman, R. M. (1994). Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: an in vivo voltammetric study. *The Journal of Neuroscience*, *14*, 442–450.

Gaspar, P., Bloch, B., & Moine, C. (1995). D1 and D2 Receptor Gene Expression in the Rat Frontal Cortex: Cellular Localization in Different Classes of Efferent Neurons. *European Journal of Neuroscience*, *7*, 1050–1063. https://doi.org/10.1111/j.1460-9568.1995.tb01092.x

Gass, N., Schwarz, A. J., Sartorius, A., Cleppien, D., Zheng, L., Schenker, E., . . . Weber-Fahr, W. (2013). Haloperidol modulates midbrain-prefrontal functional connectivity in the rat brain. *European Neuropsychopharmacology*, *23*, 1310–1319. https://doi.org/10.1016/j.euroneuro.2012.10.013

Geana, A., Wilson, R., Daw, N. D., & Cohen, J. (2016). Boredom, Information-Seeking and Exploration. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, *1*, 1751–1756.

Geddes, J. (2002). Prevention of relapse in schizophrenia. *The New England Journal of Medicine*, *346*, 56–58. https://doi.org/10.1056/NEJM200201033460112

Geddes, J., Freemantle, N., Harrison, P., & Bebbington, P. (2000). Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ*, *321*, 1371–1376.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A. (2014). *Bayesian data analysis* (3rd ed.). *Texts in statistical science*. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Applied regression and multilevel/hierarchical models*. *Analytical methods for social research*. Cambridge, New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, *5*, 189–211. https://doi.org/10.1080/19345747.2011.618213

Gelman, A., & Lokenz, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*, 457–472. https://doi.org/10.1214/ss/1177011136

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

Gether, U., Andersen, P. H., Larsson, O. M., & Schousboe, A. (2006). Neurotransmitter transporters: molecular function of important drug targets. *Trends in Pharmacological Sciences*, *27*, 375–383. https://doi.org/10.1016/j.tips.2006.05.003

Geuter, S., Qi, G., Welsh, R. C., Wager, T. D., & Lindquist, M. A. (2018). Effect Size and Power in fMRI Group Analysis. *BioRxiv.* Advance online publication. https://doi.org/10.1101/295048

Gibbs, S. E. B., & D'Esposito, M. (2005). Individual capacity differences predict working memory performance and prefrontal activity following dopamine receptor stimulation. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 212–221. https://doi.org/10.3758/CABN.5.2.212

Gibbs, S. E. B., & D'Esposito, M. (2006). A functional magnetic resonance imaging study of the effects of pergolide, a dopamine receptor agonist, on component processes of working memory. *Neuroscience*, *139*, 359–371. https://doi.org/10.1016/j.neuroscience.2005.11.055

Gill, K. M., & Grace, A. A. (2016). The Role of Neurotransmitters in Schizophrenia. In S. C. Schulz, M. F. Green, & K. J. Nelson (Eds.), *Schizophrenia and Psychotic Spectrum Disorders* (pp. 153–184). Oxford University Press. https://doi.org/10.1093/med/9780199378067.003.0010

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*, 252–269. https://doi.org/10.3758/CABN.10.2.252

Girardin, F., & Sztajzel, J. (2007). Cardiac adverse reactions associated with psychotropic drugs. *Dialogues in Clinical Neuroscience*, *9*, 92–95.

Girolami, M., & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*, 123–214. https://doi.org/10.1111/j.1467-9868.2010.00765.x

Giros, B., Jaber, M., Jones, S. R., Wightman, R. M., & Caron, M. G. (1996). Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature*, *379*, 606–612. https://doi.org/10.1038/379606a0

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B*, *41*, 148–177.

Gittins, J. C., Glazebrook, K. D., & Weber, R. (2011). *Multi-armed bandit allocation indices* (2nd ed.). Hoboken, NJ: John Wiley & Sons. Retrieved from https://doi.org/10.1002/9780470980033

Gittins, J. C., & Jones, D. M. (1974). A Dynamic Allocation Index for the Sequential Design of Experiments. In J. Gani (Ed.), *Progress in Statistics* (pp. 241–266). Amsterdam: North-Holland.

Gittins, J. C., & Whittle, P. (1989). *Multi-armed bandit allocation indices*. *Wiley-interscience series in systems and optimization*. Chichester: Wiley.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Gläscher, J., & Gitelman, D. (2008). Contrast weights in flexible factorial design with multiple groups of subjects. Retrieved from http://www.sbirc.ed.ac.uk/ cyril/download/Contrast_Weighting_Glascher_Gitelman_2008.pdf

Gläscher, J., & O'Doherty, J. (2010). Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*, 501–510. https://doi.org/10.1002/wcs.57

Glass, B. D., Maddox, W. T., Bowen, C., Savarie, Z. R., Matthews, M. D., Markman, A. B., & Schnyer, D. M. (2011). The Effects of 24-hour Sleep Deprivation on the Exploration-Exploitation Trade-off. *Biological Rhythm Research*, *42*, 99–110. https://doi.org/10.1080/09291011003726532

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 3*, 15647–15654. https://doi.org/10.1073/pnas.1014269108

Gluskin, B. S., & Mickey, B. J. (2016). Genetic variation and dopamine D2 receptor availability: a systematic review and meta-analysis of human in vivo molecular imaging studies. *Translational Psychiatry*, *6*, e747. https://doi.org/10.1038/tp.2016.22

Göbel, S., Walsh, V., & Rushworth, M. F. (2001). The mental number line and the human angular gyrus. *NeuroImage*, *14*, 1278–1289. https://doi.org/10.1006/nimg.2001.0927

Gogolla, N. (2017). The insular cortex. *Current Biology*, *27*, R580-R586. https://doi.org/10.1016/j.cub.2017.05.010

Goldman-Rakic, P. S., Muly, E. C., & Williams, G. V. (2000). D(1) receptors in prefrontal cells and circuits. *Brain Research. Brain Research Reviews*, *31*, 295–301.

Gómez-Esteban, J. C., Zarranz, J. J., Lezcano, E., Tijero, B., Luna, A., Velasco, F., . . . Garamendi, I. (2007). Influence of motor symptoms upon the quality of life of patients with Parkinson's disease. *European Neurology*, *57*, 161–165. https://doi.org/10.1159/000098468

Goto, Y., Otani, S., & Grace, A. A. (2007). The Yin and Yang of dopamine release: a new perspective. *Neuropharmacology*, *53*, 583–587. https://doi.org/10.1016/j.neuropharm.2007.07.007

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron*, *53*, 9–16. https://doi.org/10.1016/j.neuron.2006.12.009

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, *15*, 56–67. https://doi.org/10.1016/j.tics.2010.12.004

Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience*, *41*, 1–24. https://doi.org/10.1016/0306-4522(91)90196-U

Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, *95 Suppl 2*, S119-28.

Grace, A. A. (2008). Physiology of the normal and dopamine-depleted basal ganglia: insights into levodopa pharmacotherapy. *Movement Disorders*, *23 Suppl 3*, S560-9. https://doi.org/10.1002/mds.22020

Grace, A. A., & Bunney, B. S. (1983). Intracellular and extracellular electrophysiology of nigral dopaminergic neurons – 3. Evidence for electrotonic coupling. *Neuroscience*, *10*, 333–348. https://doi.org/10.1016/0306-4522(83)90137-9

Grace, A. A., & Bunney, B. S. (1984a). The control of firing pattern in nigral dopamine neurons: burst firing. *The Journal of Neuroscience*, *4*, 2877–2890.

Grace, A. A., & Bunney, B. S. (1984b). The control of firing pattern in nigral dopamine neurons: single spike firing. *The Journal of Neuroscience*, *4*, 2866–2876.

Grace, A. A., Lodge, D. J., & Buffalari, D. M. (2009). Dopamine – CNS Pathways and Neurophysiology. In *Encyclopedia of Neuroscience* (pp. 549–555). Elsevier. https://doi.org/10.1016/B978-008045046-9.01140-2

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*, 769–792. https://doi.org/10.1037/0033-2909.130.5.769

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*, 3–17. https://doi.org/10.1111/j.1469-7580.2005.00426.x

Groman, S. M., James, A. S., Seu, E., Tran, S., Clark, T. A., Harpster, S. N., . . . Jentsch, J. D. (2014). In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *The Journal of Neuroscience*, *34*, 14443–14454. https://doi.org/10.1523/JNEUROSCI.3037-14.2014

Grueschow, M., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Automatic versus Choice-Dependent Value Representations in the Human Brain. *Neuron*, *85*, 874–885. https://doi.org/10.1016/j.neuron.2014.12.054

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The Interplay Between Exploration and Exploitation. *Academy of Management Journal*, *49*, 693–706. https://doi.org/10.5465/amj.2006.22083026

Guthrie, M., Myers, C. E., & Gluck, M. A. (2009). A Neurocomputational model of tonic and phasic dopamine in action selection: A comparison with cognitive deficits in Parkinson's disease. *Behavioural Brain Research*, *200*, 48–59. https://doi.org/10.1016/j.bbr.2008.12.036

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*, 4–26. https://doi.org/10.1038/npp.2009.129

Haddad, F., Sawalha, M., Khawaja, Y., Najjar, A., & Karaman, R. (2017). Dopamine and Levodopa Prodrugs for the Treatment of Parkinson's Disease. *Molecules*, *23.* https://doi.org/10.3390/molecules23010040

Haddad, P. M., Das, A., Keyhani, S., & Chaudhry, I. B. (2012). Antipsychotic drugs and extrapyramidal side effects in first episode psychosis: a systematic review of head-head comparisons. *Journal of Psychopharmacology*, *26*, 15–26. https://doi.org/10.1177/0269881111424929

Haddad, P. M., & Dursun, S. M. (2008). Neurological complications of psychiatric drugs: clinical features and management. *Human Psychopharmacology*, *23 Suppl 1*, 15–26. https://doi.org/10.1002/hup.918

Hälbig, T., & Koller, W. (2007). Levodopa. In *Handbook of Clinical Neurology. Parkinson's Disease and Related Disorders, Part II* (Vol. 84, pp. 31–72). Elsevier. https://doi.org/10.1016/S0072-9752(07)84032-2

Hall, C. L., Humphries, M. M., & Kramer, D. L. (2007). Resource tracking by eastern chipmunks: the sampling of renewing patches. *Canadian Journal of Zoology*, *85*, 536–548. https://doi.org/10.1139/Z07-030

Hall, H., Sedvall, G., Magnusson, O., Kopp, J., Halldin, C., & Farde, L. (1994). Distribution of D1- and D2-dopamine receptors, and dopamine and its metabolites in the human brain. *Neuropsychopharmacology*, *11*, 245–256. https://doi.org/10.1038/sj.npp.1380111

Hammond, C., Bergman, H., & Brown, P. (2007). Pathological synchronization in Parkinson's disease: networks, models and treatments. *Trends in Neurosciences*, *30*, 357–364. https://doi.org/10.1016/j.tins.2007.05.004

Hampton, A. N., & O'Doherty, J. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1377–1382. https://doi.org/10.1073/pnas.0606297104

Handley, R., Zelaya, F. O., Reinders, A. A. T. S., Marques, T. R., Mehta, M. A., O'Gorman, R., . . . Dazzan, P. (2013). Acute effects of single-dose aripiprazole and haloperidol on resting cerebral blood flow (rCBF) in the human brain. *Human Brain Mapping*, *34*, 272–282. https://doi.org/10.1002/hbm.21436

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience*, *28*, 5623–5630. https://doi.org/10.1523/JNEUROSCI.1309-08.2008

Hariri, A. R., Brown, S. M., Williamson, D. E., Flory, J. D., de Wit, H., & Manuck, S. B. (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *The Journal of Neuroscience*, *26*, 13213–13217. https://doi.org/10.1523/JNEUROSCI.3446-06.2006

Harlé, K. M., Zhang, S., Schiff, M., Mackey, S., Paulus, M. P., & Yu, A. J. (2015). Altered Statistical Learning and Decision-Making in Methamphetamine Dependence: Evidence from a Two-Armed Bandit Task. *Frontiers in Psychology*, *6*, 1910. https://doi.org/10.3389/fpsyg.2015.01910

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. M. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *The Journal of Neuroscience*, *34*, 698–704. https://doi.org/10.1523/JNEUROSCI.2489-13.2014

Hartley, C. A., & Phelps, E. A. (2012). Anxiety and decision-making. *Biological Psychiatry*, *72*, 113–118. https://doi.org/10.1016/j.biopsych.2011.12.027

Harun, R., Hare, K. M., Brough, E. M., Munoz, M. J., Grassi, C. M., Torres, G. E., . . . Wagner, A. K. (2016). Fast-scan cyclic voltammetry demonstrates that L-DOPA produces dose-dependent, regionally selective bimodal effects on striatal dopamine kinetics in vivo. *Journal of Neurochemistry*, *136*, 1270–1283. https://doi.org/10.1111/jnc.13444

Hauser, R. A. (2009). Levodopa: past, present, and future. *European Neurology*, *62*, 1–8. https://doi.org/10.1159/000215875

Hayden, B. Y., Nair, A. C., McCoy, A. N., & Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron*, *60*, 19–25. https://doi.org/10.1016/j.neuron.2008.09.012

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*, 933-939. https://doi.org/10.1038/nn.2856

Hayes, T. R., & Petrov, A. A. (2016). Pupil Diameter Tracks the Exploration-Exploitation Trade-off during Analogical Reasoning and Explains Individual Differences in Fluid Intelligence. *Journal of Cognitive Neuroscience*, *28*, 308–318. https://doi.org/10.1162/jocn_a_00895

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews. Neuroscience*, *3*, 142–151. https://doi.org/10.1038/nrn730

Hefti, F., & Melamed, E. (1980). L-DOPA's mechanism of action in Parkinson's disease. *Trends in Neurosciences*, *3*, 229–231. https://doi.org/10.1016/S0166-2236(80)80070-1

Hellwege, J., Keaton, J., Giri, A., Gao, X., Velez Edwards, D. R., & Edwards, T. L. (2017). Population Stratification in Genetic Association Studies. *Current Protocols in Human Genetics*, *95*, 1.22.1-1.22.23. https://doi.org/10.1002/cphg.48

Helversen, B. von, Mata, R., Samanez-Larkin, G. R., & Wilke, A. (2018). Foraging, exploration, or search? On the (lack of) convergent validity between three behavioral paradigms. *Evolutionary Behavioral Sciences*, *12*, 152–162. https://doi.org/10.1037/ebs0000121

Henson, R. N., & Friston, K. J. (2006). Convolution Models for fMRI. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping* (pp. 178–192). San Diego: Academic Press.

Henson, R. N., & Penny, W. D. (2005). *ANOVAs and SPM* (Technical Report). London: Institute of Cognitive Neuroscience, Wellcome Department of Imaging Neuroscience. Retrieved from https://www.fil.ion.ucl.ac.uk/~wpenny/publications/rik_anova.pdf

Hernandez, L., & Hoebel, B. G. (1989). Haloperidol given chronically decreases basal dopamine in the prefrontal cortex more than the striatum or nucleus accumbens as simultaneously measured by microdialysis. *Brain Research Bulletin*, *22*, 763–769. https://doi.org/10.1016/0361-9230(89)90097-X

Hershey, T. (2003). Long term treatment and disease severity change brain responses to levodopa in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*, 844–851. https://doi.org/10.1136/jnnp.74.7.844

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*, 46–54. https://doi.org/10.1016/j.tics.2014.10.004

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*, 304–320. https://doi.org/10.1037/a0026767

Hirvonen, M., Laakso, A., Någren, K., Rinne, J. O., Pohjalainen, T., & Hietala, J. (2004). C(957)T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability in vivo. *Molecular Psychiatry*, *9*, 1060–1061. https://doi.org/10.1038/sj.mp.4001561

Hökfelt, T., Fuxe, K., & Goldstein, M. (1973). Immunohistochemical localization of aromatic L-amino acid decarboxylase (DOPA decarboxylase) in central dopamine and 5-hydroxytryptamine nerve cell bodies of the rat. *Brain Research*, *53*, 175–180. https://doi.org/10.1016/0006-8993(73)90776-2

Honey, G. D., Suckling, J., Zelaya, F., Long, C., Routledge, C., Jackson, S., . . . Bullmore, E. T. (2003). Dopaminergic drug effects on physiological connectivity in a human cortico-striato-thalamic system. *Brain*, *126*, 1767–1781. https://doi.org/10.1093/brain/awg184

Horne, M. K., Cheng, C. H., & Wooten, G. F. (1984). The cerebral metabolism of L-dihydroxyphenylalanine. An autoradiographic and biochemical study. *Pharmacology*, *28*, 12–26. https://doi.org/10.1159/000137938

Hornykiewicz, O. (1966). Dopamine (3-hydroxytyramine) and brain function. *Pharmacological Reviews*, *18*, 925–964.

Hornykiewicz, O. (2002). L-DOPA: from a biologically inactive amino acid to a successful therapeutic agent. *Amino Acids*, *23*, 65–70. https://doi.org/10.1007/s00726-001-0111-9

Hornykiewicz, O. (2017). L-DOPA. *Journal of Parkinson's Disease*, *7*, S3-S10. https://doi.org/10.3233/JPD-179004

Housden, C. R., O'Sullivan, S. S., Joyce, E. M., Lees, A. J., & Roiser, J. P. (2010). Intact reward learning but elevated delay discounting in Parkinson's disease patients with impulsive-compulsive spectrum behaviors. *Neuropsychopharmacology*, *35*, 2155–2164. https://doi.org/10.1038/npp.2010.84

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III – the final common pathway. *Schizophrenia Bulletin*, *35*, 549–562. https://doi.org/10.1093/schbul/sbp006

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science*, *310*, 1680–1683. https://doi.org/10.1126/science.1115327

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *The Journal of Neuroscience*, *25*, 3304–3311. https://doi.org/10.1523/JNEUROSCI.5070-04.2005

Humphreys, K. L., Lee, S. S., Telzer, E. H., Gabard-Durnam, L. J., Goff, B., Flannery, J., & Tottenham, N. (2015). Exploration-exploitation strategy is dependent on early experience. *Developmental Psychobiology*, *57*, 313–321. https://doi.org/10.1002/dev.21293

Humphries, M. D., Khamassi, M., & Gurney, K. (2012). Dopaminergic Control of the Exploration-Exploitation Trade-Off via the Basal Ganglia. *Frontiers in Neuroscience*, *6*, 9. https://doi.org/10.3389/fnins.2012.00009

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, *15*, 470-6, S1-3. https://doi.org/10.1038/nn.3017

Huot, P., Fox, S. H., & Brotchie, J. M. (2016). Dopamine Reuptake Inhibitors in Parkinson's Disease: A Review of Nonhuman Primate Studies and Clinical Trials. *The Journal of Pharmacology and Experimental Therapeutics*, *357*, 562–569. https://doi.org/10.1124/jpet.116.232371

Hurd, Y. L., Suzuki, M., & Sedvall, G. C. (2001). D1 and D2 dopamine receptor mRNA expression in whole hemisphere sections of the human brain. *Journal of Chemical Neuroanatomy*, *22*, 127–137. https://doi.org/10.1016/S0891-0618(01)00122-3

Hutchinson, J. M.C., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour*, *75*, 1331–1349. https://doi.org/10.1016/j.anbehav.2007.09.006

Hyland, B. I., Reynolds, J. N. J., Hay, J., Perk, C. G., & Miller, R. (2002). Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience*, *114*, 475–492.

Ishii, S., Yoshida, W., & Yoshimoto, J. (2002). Control of exploitation–exploration meta-parameter in reinforcement learning. *Neural Networks*, *15*, 665–687. https://doi.org/10.1016/S0893-6080(02)00056-4

Iversen, L. L. (Ed.). (2010). *Dopamine handbook*. Oxford: Oxford Univ. Press.

Iwaki, H., Nishikawa, N., Nagai, M., Tsujii, T., Yabe, H., Kubo, M., . . . Nomoto, M. (2015). Pharmacokinetics of levodopa/benserazide versus levodopa/carbidopa in healthy subjects and patients with Parkinson's disease. *Neurology and Clinical Neuroscience*, *3*, 68–73. https://doi.org/10.1111/ncn3.152

Jankovic, J. (2005). Motor fluctuations and dyskinesias in Parkinson's disease: clinical manifestations. *Movement Disorders*, *20 Suppl 11*, S11-6. https://doi.org/10.1002/mds.20458

Janssen, P. A. J., van de Westeringh, C., Jageneau, A. H. M., Demoen, P. J. A., Hermans, B. K. F., van Daele, G. H. P. V., . . . Niemegeers, C. J. E. (1959). Chemistry and Pharmacology of CNS Depressants Related to 4-(4-Hydroxy-4-phenylpiperidino)butyrophenone Part I – Synthesis and screening data in mice. *Journal of Medicinal and Pharmaceutical Chemistry*, *1*, 281–297.

Jaworski, J. N., Gonzales, R. A., & Randall, P. K. (2001). Effect of dopamine D2/D3 receptor antagonist sulpiride on amphetamine-induced changes in striatal extracellular dopamine. *European Journal of Pharmacology*, *418*, 201–206. https://doi.org/10.1016/S0014-2999(01)00936-0

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*, 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Jepma, M., Te Beek, E. T., Wagenmakers, E.-J., van Gerven, J. M. A., & Nieuwenhuis, S. (2010). The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Frontiers in Human Neuroscience*, *4*, 170. https://doi.org/10.3389/fnhum.2010.00170

Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *The Journal of Neuroscience*, *31*, 1606–1613. https://doi.org/10.1523/JNEUROSCI.3904-10.2011

Jones, C. L., Minati, L., Harrison, N. A., Ward, J., & Critchley, H. D. (2011). Under pressure: response urgency modulates striatal and insula activity during decision-making under risk. *PloS One*, *6*, e20942. https://doi.org/10.1371/journal.pone.0020942

Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function – A review. *Neuroscience and Biobehavioral Reviews*, *71*, 58–82. https://doi.org/10.1016/j.neubiorev.2016.08.020

Jönsson, E. G., Nöthen, M. M., Grünhage, F., Farde, L., Nakashima, Y., Propping, P., & Sedvall, G. C. (1999). Polymorphisms in the dopamine D2 receptor gene and their relationships to striatal dopamine receptor density of healthy volunteers. *Molecular Psychiatry*, *4*, 290–296. https://doi.org/10.1038/sj.mp.4000532

Jordan, S., Eberling, J. L., Bankiewicz, K. S., Rosenberg, D., Coxson, P. G., VanBrocklin, H. F., . . . Jagust, W. J. (1997). 6-[18F]Fluoro-l-m-tyrosine: metabolism, positron emission tomography kinetics, and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine lesions in primates. *Brain Research*, *750*, 264–276. https://doi.org/10.1016/S0006-8993(96)01366-2

Kaakkola, S. (2000). Clinical Pharmacology, Therapeutic Use and Potential of COMT Inhibitors in Parkinson's Disease. *Drugs*, *59*, 1233–1250. https://doi.org/10.2165/00003495-200059060-00004

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*, 1625–1633. https://doi.org/10.1038/nn2007

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, *63*, 733–745. https://doi.org/10.1016/j.neuron.2009.09.003

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Käenmäki, M., Tammimäki, A., Myöhänen, T., Pakarinen, K., Amberg, C., Karayiorgou, M., . . . Männistö, P. T. (2010). Quantitative role of COMT in dopamine clearance in the prefrontal cortex of freely moving mice. *Journal of Neurochemistry*, *114*, 1745–1755. https://doi.org/10.1111/j.1471-4159.2010.06889.x

Kahnt, T., & Tobler, P. N. (2017). Dopamine Modulates the Functional Organization of the Orbitofrontal Cortex. *The Journal of Neuroscience*, *37*, 1493–1504. https://doi.org/10.1523/JNEUROSCI.2827-16.2016

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition*, *29*, 1–9. https://doi.org/10.3758/BF03195735

Kakade, S., & Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Networks*, *15*, 549–559. https://doi.org/10.1016/S0893-6080(02)00048-5

Kalant, H. (2001). The pharmacology and toxicology of "ecstasy" (MDMA) and related drugs. *Canadian Medical Association Journal*, *165*, 917–928.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*, 35. https://doi.org/10.1115/1.3662552

Kalman, R. E., & Bucy, R. S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, *83*, 95. https://doi.org/10.1115/1.3658902

Kaminer, J., Powers, A. S., Horn, K. G., Hui, C., & Evinger, C. (2011). Characterizing the spontaneous blink generator: an animal model. *The Journal of Neuroscience*, *31*, 11256–11267. https://doi.org/10.1523/JNEUROSCI.6218-10.2011

Kaminer, J., Thakur, P., & Evinger, C. (2015). Effects of subthalamic deep brain stimulation on blink abnormalities of 6-OHDA lesioned rats. *Journal of Neurophysiology*, *113*, 3038–3046. https://doi.org/10.1152/jn.01072.2014

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*, 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kapur, S., Agid, O., Mizrahi, R., & Li, M. (2006). How antipsychotics work – From receptors to reality. *NeuroRx*, *3*, 10–21. https://doi.org/10.1016/j.nurx.2005.12.003

Kapur, S., Zipursky, R., Jones, C., Remington, G., & Houle, S. (2000). Relationship between dopamine D(2) occupancy, clinical response, and side effects: a double-blind PET study of first-episode schizophrenia. *The American Journal of Psychiatry*, *157*, 514–520. https://doi.org/10.1176/appi.ajp.157.4.514

Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, *106*, 643–653.

Karson, C. N., Bigelow, L. B., Kleinman, J. E., Weinberger, D. R., & Wyatt, R. J. (1982). Haloperidol-induced changes in blink rates correlate with changes in BPRS score. *British Journal of Psychiatry*, *140*, 503–507.

Karson, C. N., Freed, W. J., Kleinman, J. E., Bigelow, L. B., & Wyatt, R. J. (1981). Neuroleptics decrease blinking in schizophrenic subjects. *Biological Psychiatry*, *16*, 679–682.

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58. https://doi.org/10.1016/j.jmp.2016.03.007

Katz, K., & Naug, D. (2015). Energetic state regulates the exploration–exploitation trade-off in honeybees. *Behavioral Ecology*, *26*, 1045–1050. https://doi.org/10.1093/beheco/arv045

Katz, P. S., & Calin-Jageman, R. J. (2009). Neuromodulation. In *Encyclopedia of Neuroscience* (pp. 497–503). Elsevier. https://doi.org/10.1016/B978-008045046-9.01964-1

Katzenschlager, R., & Lees, A. J. (2002). Treatment of Parkinson's disease: levodopa as the first choice. *Journal of Neurology*, *249 Suppl 2*, II19-24. https://doi.org/10.1007/s00415-002-1204-4

Kayser, A. S., Mitchell, J. M., Weinstein, D., & Frank, M. J. (2015). Dopamine, locus of control, and the exploration-exploitation tradeoff. *Neuropsychopharmacology*, *40*, 454–462. https://doi.org/10.1038/npp.2014.193

Keeler, J. F., Pretsell, D. O., & Robbins, T. W. (2014). Functional implications of dopamine D1 vs. D2 receptors: A 'prepare and select' model of the striatal direct vs. indirect pathways. *Neuroscience*, *282*, 156–175. https://doi.org/10.1016/j.neuroscience.2014.07.021

Kehr, J., & Yoshitake, T. (2013). Monitoring molecules in neuroscience: historical overview and current advancements. *Frontiers in Bioscience (Elite Edition)*, *5*, 947–954.

Kellendonk, C., Simpson, E. H., Polan, H. J., Malleret, G., Vronskaya, S., Winiger, V., . . . Kandel, E. (2006). Transient and selective overexpression of dopamine D2 receptors in the striatum causes persistent abnormalities in prefrontal cortex functioning. *Neuron*, *49*, 603–615. https://doi.org/10.1016/j.neuron.2006.01.023

Keller, G. A., Czerniuk, P., Bertuola, R., Spatz, J. G., Assefi, A. R., & Di Girolamo, G. (2011). Comparative bioavailability of 2 tablet formulations of levodopa/benserazide in healthy, fasting volunteers: a single-dose, randomized-sequence, open-label crossover study. *Clinical Therapeutics*, *33*, 500–510. https://doi.org/10.1016/j.clinthera.2011.04.012

Keller, R. W., Kuhr, W. G., Wightman, R. M., & Zigmond, M. J. (1988). The effect of L-DOPA on in vivo dopamine release from nigrostriatal bundle neurons. *Brain Research*, *447*, 191–194. https://doi.org/10.1016/0006-8993(88)90985-7

Kennedy, R. T., Jones, S. R., & Wightman, R. M. (1992). Dynamic observation of dopamine autoreceptor effects in rat striatal slices. *Journal of Neurochemistry*, *59*, 449–455.

Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., & Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, *9*, 940–947. https://doi.org/10.1038/nn1724

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026. https://doi.org/10.1126/science.1089910

Kessels, R. P., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment*, *15*, 426–434. https://doi.org/10.1177/1073191108315611

Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., & de Haan, E. H. (2000). The Corsi Block-Tapping Task: standardization and normative data. *Applied Neuropsychology*, *7*, 252–258. https://doi.org/10.1207/S15324826AN0704_8

Khor, S.-P., & Hsu, A. (2007). The pharmacokinetics and pharmacodynamics of levodopa in the treatment of Parkinson's disease. *Current Clinical Pharmacology*, *2*, 234–243.

Kiebel, S. J., & Holmes, A. P. (2006). The General Linear Model. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping.* San Diego: Academic Press.

Kim, K. M., Baratta, M. V., Yang, A., Lee, D., Boyden, E. S., & Fiorillo, C. D. (2012). Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. *PloS One*, *7*, e33612. https://doi.org/10.1371/journal.pone.0033612

Kimber, T. E., & Thompson, P. D. (2000). Increased blink rate in advanced Parkinson's disease: A form of 'off'-period dystonia? *Movement Disorders*, *15*, 982–985. https://doi.org/10.1002/1531-8257(200009)15:5<982::AID-MDS1033>3.0.CO;2-P

Kimberg, D. Y., & D'Esposito, M. (2003). Cognitive effects of the dopamine receptor agonist pergolide. *Neuropsychologia*, *41*, 1020–1027. https://doi.org/10.1016/S0028-3932(02)00317-2

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *NeuroReport*, *8*, 3581–3585. https://doi.org/10.1097/00001756-199711100-00032

Kirsch, P., Reuter, M., Mier, D., Lonsdorf, T., Stark, R., Gallhofer, B., . . . Hennig, J. (2006). Imaging gene-substance interactions: the effect of the DRD2 TaqIA polymorphism and the dopamine agonist bromocriptine on the brain activation during the anticipation of reward. *Neuroscience Letters*, *405*, 196–201. https://doi.org/10.1016/j.neulet.2006.07.030

Kish, S. J., Shannak, K., & Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. *The New England Journal of Medicine*, *318*, 876–880. https://doi.org/10.1056/NEJM198804073181402

Kitahama, K., Ikemoto, K., Jouvet, A., Araneda, S., Nagatsu, I., Raynaud, B., . . . Niwa, S.-I. (2009). Aromatic L-amino acid decarboxylase-immunoreactive structures in human midbrain, pons, and medulla. *Journal of Chemical Neuroanatomy*, *38*, 130–140. https://doi.org/10.1016/j.jchemneu.2009.06.010

Klanker, M., Feenstra, M., & Denys, D. (2013). Dopaminergic control of cognitive flexibility in humans and animals. *Frontiers in Neuroscience*, *7*, 201. https://doi.org/10.3389/fnins.2013.00201

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*, 1–16.

Kleinman, J. E., Karson, C. N., Weinberger, D. R., Freed, W. J., Berman, K. F., & Wyatt, R. J. (1984). Eye-blinking and cerebral ventricular size in chronic schizophrenic patients. *The American Journal of Psychiatry*, *141*, 1430–1432. https://doi.org/10.1176/ajp.141.11.1430

Knecht, S., Breitenstein, C., Bushuven, S., Wailke, S., Kamping, S., Flöel, A., . . . Ringelstein, E. B. (2004). Levodopa: faster and better word learning in normal humans. *Annals of Neurology*, *56*, 20–26. https://doi.org/10.1002/ana.20125

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, *2: 398.* https://doi.org/10.3389/fpsyg.2011.00398

Knutson, B., & Gibbs, S. E. B. (2007). Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology*, *191*, 813–822. https://doi.org/10.1007/s00213-006-0686-7

Kobayashi, S., & Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. *The Journal of Neuroscience*, *28*, 7837–7846. https://doi.org/10.1523/JNEUROSCI.1600-08.2008

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*, 594–598. https://doi.org/10.1126/science.1142995

Kohno, M., Ghahremani, D. G., Morales, A. M., Robertson, C. L., Ishibashi, K., Morgan, A. T., . . . London, E. D. (2015). Risk-taking behavior: dopamine D2/D3 receptors, feedback, and frontolimbic activity. *Cerebral Cortex*, *25*, 236–245. https://doi.org/10.1093/cercor/bht218

Kohno, M., Nurmi, E. L., Laughlin, C. P., Morales, A. M., Gail, E. H., Hellemann, G. S., & London, E. D. (2016). Functional Genetic Variation in Dopamine Signaling Moderates Prefrontal Cortical Activity During Risky Decision Making. *Neuropsychopharmacology*, *41*, 695–703. https://doi.org/10.1038/npp.2015.192

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, *336*, 95–98. https://doi.org/10.1126/science.1216930

Kolling, N., Behrens, T. E. J., Wittmann, M. K., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology*, *37*, 36–43. https://doi.org/10.1016/j.conb.2015.12.007

Kömek, K., Bard Ermentrout, G., Walker, C. P., & Cho, R. Y. (2012). Dopamine and gamma band synchrony in schizophrenia – insights from computational and empirical studies. *The European Journal of Neuroscience*, *36*, 2146–2155. https://doi.org/10.1111/j.1460-9568.2012.08071.x

Korchounov, A., Meyer, M. F., & Krasnianski, M. (2010). Postsynaptic nigrostriatal dopamine receptors and their role in movement regulation. *Journal of Neural Transmission*, *117*, 1359–1369. https://doi.org/10.1007/s00702-010-0454-z

Kotani, M., Kiyoshi, A., Murai, T., Nakako, T., Matsumoto, K., Matsumoto, A., . . . Ikeda, K. (2016). The dopamine D1 receptor agonist SKF-82958 effectively increases eye blinking count in common marmosets. *Behavioural Brain Research*, *300*, 25–30. https://doi.org/10.1016/j.bbr.2015.11.028

Kovach, C. K., Daw, N. D., Rudrauf, D., Tranel, D., O'Doherty, J., & Adolphs, R. (2012). Anterior prefrontal cortex contributes to action selection through tracking of recent reward trends. *The Journal of Neuroscience*, *32*, 8434–8442. https://doi.org/10.1523/JNEUROSCI.5468-11.2012

Kravitz, A. V., Tye, L. D., & Kreitzer, A. C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neuroscience*, *15*, 816–818. https://doi.org/10.1038/nn.3100

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*, 341–372. https://doi.org/10.1016/j.pneurobio.2004.03.006

Kroemer, N. B., Lee, Y., Pooseh, S., Eppinger, B., Goschke, T., & Smolka, M. N. (2018). L-DOPA reduces model-free control of behavior by attenuating the transfer of value to action. *NeuroImage*, *186*, 113–125. https://doi.org/10.1016/j.neuroimage.2018.10.075

Krueger, P. M., Wilson, R. C., & Cohen, J. D. (2017). Strategies for exploration in the domain of losses. *Judgment and Decision Making*, *12*, 104–117.

Krugel, L. K., Biele, G., Mohr, P. N. C., Li, S.-C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 17951–17956. https://doi.org/10.1073/pnas.0905191106

Kruis, A., Slagter, H. A., Bachhuber, D. R. W., Davidson, R. J., & Lutz, A. (2016). Effects of meditation practice on spontaneous eyeblink rate. *Psychophysiology*, *53*, 749–758. https://doi.org/10.1111/psyp.12619

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology. General*, *142*, 573–603. https://doi.org/10.1037/a0029146

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Amsterdam: Academic Press.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian Estimation in Hierarchical Models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279–299). Oxford: Oxford University Press.

Kudo, S., & Ishizaki, T. (1999). Pharmacokinetics of haloperidol: an update. *Clinical Pharmacokinetics*, *37*, 435–456. https://doi.org/10.2165/00003088-199937060-00001

Kunishio, K., & Haber, S. N. (1994). Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. *The Journal of Comparative Neurology*, *350*, 337–356. https://doi.org/10.1002/cne.903500302

Kuroki, T., Meltzer, H. Y., & Ichikawa, J. (1999). Effects of antipsychotic drugs on extracellular dopamine levels in rat medial prefrontal cortex and nucleus accumbens. *The Journal of Pharmacology and Experimental Therapeutics*, *288*, 774–781.

Labbate, L. A. (2010). *Handbook of psychiatric drug therapy* (6th ed.). Philadelphia [etc.]: Wolters Kluwer/Lippincott Williams & Wilkins.

Lacerda, A., Santos, R. L. T., Veloso, A., & Ziviani, N. (2015). Improving daily deals recommendation using explore-then-exploit strategies. *Information Retrieval Journal*, *18*, 95–122. https://doi.org/10.1007/s10791-014-9249-4

Lai, T.L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, *6*, 4–22. https://doi.org/10.1016/0196-8858(85)90002-8

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., & Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron*, *57*, 760–773. https://doi.org/10.1016/j.neuron.2008.01.022

Lancaster, T. M., Linden, D. E., & Heerey, E. A. (2012). COMT val158met predicts reward responsiveness in humans. *Genes, Brain, and Behavior*, *11*, 986–992. https://doi.org/10.1111/j.1601-183X.2012.00838.x

Landau, S. M., Lal, R., O'Neil, J. P., Baker, S., & Jagust, W. J. (2009). Striatal dopamine and working memory. *Cerebral Cortex*, *19*, 445–454. https://doi.org/10.1093/cercor/bhn095

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. *Technology in Mental Health Care Delivery Systems*, 119–137.

Laruelle, M. (2000). Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. *Journal of Cerebral Blood Flow and Metabolism*, *20*, 423–451. https://doi.org/10.1097/00004647-200003000-00001

Latty, T., & Beekman, M. (2013). Keeping track of changes: the performance of ant colonies in dynamic environments. *Animal Behaviour*, *85*, 637–643. https://doi.org/10.1016/j.anbehav.2012.12.027

Lau, B., & Glimcher, P. W. (2005). Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *Journal of the Experimental Analysis of Behavior*, *84*, 555–579. https://doi.org/10.1901/jeab.2005.110-04

Lau, C.-I., Wang, H.-C., Hsu, J.-L., & Liu, M.-E. (2013). Does the dopamine hypothesis explain schizophrenia? *Reviews in the Neurosciences*, *24*, 389–400. https://doi.org/10.1515/revneuro-2013-0011

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration-exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*, 319–338. https://doi.org/10.1002/smj.2221

Laureiro-Martínez, D., Canessa, N., Brusoni, S., Zollo, M., Hare, T., Alemanno, F., & Cappa, S. F. (2014). Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration-exploitation task. *Frontiers in Human Neuroscience*, *7: 927*, 1–10. https://doi.org/10.3389/fnhum.2013.00927

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and Exploitation Within and Across Organizations. *Academy of Management Annals*, *4*, 109–155. https://doi.org/10.5465/19416521003691287

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, *64*, 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Lee, B., London, E. D., Poldrack, R. A., Farahi, J., Nacca, A., Monterosso, J. R., . . . Mandelkern, M. A. (2009). Striatal dopamine d2/d3 receptor availability is reduced in methamphetamine dependence and is linked to impulsivity. *The Journal of Neuroscience*, *29*, 14734–14740. https://doi.org/10.1523/JNEUROSCI.3765-09.2009

Lee, M. D., & Wagenmakers, E.-J. (2015). *Bayesian cognitive modeling: A practical course* (Repr). Cambridge: Cambridge University Press.

Lee, M. D., Zhang, S., Munro, M., & Steyvers, M. (2011). Psychological models of human and optimal performance in bandit problems. *Cognitive Systems Research*, *12*, 164–174. https://doi.org/10.1016/j.cogsys.2010.07.007

Leh, S. E., Petrides, M., & Strafella, A. P. (2010). The neural circuitry of executive functions in healthy subjects and Parkinson's disease. *Neuropsychopharmacology*, *35*, 70–85. https://doi.org/10.1038/npp.2009.88

Lehmann, E., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York, NY: Springer.

Lejuez, C. W., Aklin, W., Bornovalova, M., & Moolchan, E. T. (2005). Differences in risk-taking propensity across inner-city adolescent ever- and never-smokers. *Nicotine & Tobacco Research*, *7*, 71–79. https://doi.org/10.1080/14622200412331328484

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology. Applied*, *8*, 75–84.

Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017). Chronic and Acute Stress Promote Overexploitation in Serial Decision Making. *The Journal of Neuroscience*, *37*, 5681–5689. https://doi.org/10.1523/JNEUROSCI.3618-16.2017

Leonard, C. E., Freeman, C. P., Newcomb, C. W., Bilker, W. B., Kimmel, S. E., Strom, B. L., & Hennessy, S. (2013). Antipsychotics and the Risks of Sudden Cardiac Death and All-Cause Death: Cohort Studies in Medicaid and Dually-Eligible Medicaid-Medicare Beneficiaries of Five States. *Journal of Clinical & Experimental Cardiology*, *Suppl 10*, 1–9. https://doi.org/10.4172/2155-9880.S10-006

Lerner, A., Bagic, A., Hanakawa, T., Boudreau, E. A., Pagan, F., Mari, Z., . . . Hallett, M. (2009). Involvement of insula and cingulate cortices in control and suppression of natural urges. *Cerebral Cortex*, *19*, 218–223. https://doi.org/10.1093/cercor/bhn074

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Örey, D., Richter, F., . . . Davis, J. M. (2013). Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *The Lancet*, *382*, 951–962. https://doi.org/10.1016/S0140-6736(13)60733-3

Leucht, S., Corves, C., Arbter, D., Engel, R. R., Li, C., & Davis, J. M. (2009). Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *The Lancet*, *373*, 31–41. https://doi.org/10.1016/S0140-6736(08)61764-X

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, *22*, 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Levy, F. (2009). Dopamine vs noradrenaline: inverted-U effects and ADHD theories. *The Australian and New Zealand Journal of Psychiatry*, *43*, 101–108. https://doi.org/10.1080/00048670802607238

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology*, *103*, 1036–1047. https://doi.org/10.1152/jn.00853.2009

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage Publications.

Lewis, D. A., Melchitzky, D. S., Sesack, S. R., Whitehead, R. E., Auh, S., & Sampson, A. (2001). Dopamine transporter immunoreactivity in monkey cerebral cortex: Regional, laminar, and ultrastructural localization. *The Journal of Comparative Neurology*, *432*, 119–136. https://doi.org/10.1002/cne.1092

Li, P., Snyder, G. L., & Vanover, K. E. (2016). Dopamine Targeting Drugs for the Treatment of Schizophrenia: Past, Present and Future. *Current Topics in Medicinal Chemistry*, *16*, 3385–3403. https://doi.org/10.2174/1568026616666160608084834

Li, Y., Tesson, B. M., Churchill, G. A., & Jansen, R. C. (2010). Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends in Genetics : TIG*, *26*, 493–498. https://doi.org/10.1016/j.tig.2010.09.002

Lidow, M. S., Goldman-Rakic, P. S., Gallager, D. W., & Rakic, P. (1991). Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using 3Hraclopride, 3Hspiperone and 3HSCH23390. *Neuroscience*, *40*, 657–671.

LimeSurvey GmbH. *LimeSurvey: An Open Source survey tool*. Hamburg, Germany. Retrieved from http://www.limesurvey.org

Linnet, J., Mouridsen, K., Peterson, E., Møller, A., Doudet, D. J., & Gjedde, A. (2012). Striatal dopamine release codes uncertainty in pathological gambling. *Psychiatry Research*, *204*, 55–60. https://doi.org/10.1016/j.pscychresns.2012.04.012

Litvan, I., Bhatia, K. P., Burn, D. J., Goetz, C. G., Lang, A. E., McKeith, I., . . . Wenning, G. K. (2003). Movement Disorders Society Scientific Issues Committee report: SIC Task Force appraisal of clinical diagnostic criteria for Parkinsonian disorders. *Movement Disorders*, *18*, 467–486. https://doi.org/10.1002/mds.10459

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*, 145–163. https://doi.org/10.1152/jn.1992.67.1.145

Lloyd, K. G., Davidson, L., & Hornykiewicz, O. (1975). The neurochemistry of Parkinson's disease: effect of L-DOPA therapy. *The Journal of Pharmacology and Experimental Therapeutics*, *195*, 453–464.

Lloyd, K. G., & Hornykiewicz, O. (1972). Occurrence and Distribution of Aromatic L-Amino Acid (L-DOPA) Decarboxylase in the Human Brain. *Journal of Neurochemistry*, *19*, 1549–1559. https://doi.org/10.1111/j.1471-4159.1972.tb05099.x

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*, 1003–1037. https://doi.org/10.1098/rstb.2002.1114

López-Avila, A., Coffeen, U., Ortega-Legaspi, J. M., del Angel, R., & Pellicer, F. (2004). Dopamine and NMDA systems modulate long-term nociception in the rat anterior cingulate cortex. *Pain*, *111*, 136–143. https://doi.org/10.1016/j.pain.2004.06.010

López-Muñoz, F., & Alamo, C. (2009). The consolidation of neuroleptic therapy: Janssen, the discovery of haloperidol and its introduction into clinical practice. *Brain Research Bulletin*, *79*, 130–141. https://doi.org/10.1016/j.brainresbull.2009.01.005

Lorist, M. M., Boksem, M. A. S., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Brain Research. Cognitive Brain Research*, *24*, 199–205. https://doi.org/10.1016/j.cogbrainres.2005.01.018

Luciana, M. (1998). Opposing roles for dopamine and serotonin in the modulation of human spatial working memory functions. *Cerebral Cortex*, *8*, 218–226. https://doi.org/10.1093/cercor/8.3.218

Lüscher, C., & Malenka, R. C. (2011). Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. *Neuron*, *69*, 650–663. https://doi.org/10.1016/j.neuron.2011.01.017

Lüscher, C., & Ungless, M. A. (2006). The mechanistic classification of addictive drugs. *PLoS Medicine*, *3*, e437. https://doi.org/10.1371/journal.pmed.0030437

Mackert, A., Woyth, C., Flechtner, M., & Frick, K. (1988). Increased blink rate in acute and remitted schizophrenics. *Pharmacopsychiatry*, *21*, 334–335. https://doi.org/10.1055/s-2007-1016999

Macready, W. G., & Wolpert, D. H. (1998). Bandit problems and the exploration/exploitation tradeoff. *IEEE Transactions on Evolutionary Computation*, *2*, 2–22. https://doi.org/10.1109/4235.728210

Mahajan, D. K., Rastogi, R., Tiwari, C., & Mitra, A. (2012). Log UCB: an explore-exploit algorithm for comments recommendation. In X. Chen (Ed.), *Proceedings of the 21st ACM international conference on Information and knowledge management* (p. 6). New York, NY: ACM. https://doi.org/10.1145/2396761.2396767

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel Pattern Analysis for fMRI Data: A Review. *Computational and Mathematical Methods in Medicine*, *2012*. https://doi.org/10.1155/2012/961257

Malhotra, A. K., Kestler, L. J., Mazzanti, C., Bates, J. A., Goldberg, T., & Goldman, D. (2002). A functional polymorphism in the COMT gene and performance on a test of prefrontal cognition. *The American Journal of Psychiatry*, *159*, 652–654. https://doi.org/10.1176/appi.ajp.159.4.652

Mandali, A., Rengaswamy, M., Chakravarthy, V. S., & Moustafa, A. A. (2015). A spiking Basal Ganglia model of synchrony, exploration and decision making. *Frontiers in Neuroscience*, *9*, 191. https://doi.org/10.3389/fnins.2015.00191

Männistö, P. T., & Kaakkola, S. (1999). Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacological Reviews*, *51*, 593–628.

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, *10*, L55-L58. https://doi.org/10.1088/0022-3719/10/3/004

Mansouri, F. A., Buckley, M. J., Mahboubi, M., & Tanaka, K. (2015). Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, E3940-E3949. https://doi.org/10.1073/pnas.1422629112

Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals – a key role for the frontopolar cortex. *Nature Reviews. Neuroscience*, *18*, 645–657. https://doi.org/10.1038/nrn.2017.111

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*, 71–87. https://doi.org/10.1287/orsc.2.1.71

Marder, E. (2012). Neuromodulation of neuronal circuits: back to the future. *Neuron*, *76*, 1–11. https://doi.org/10.1016/j.neuron.2012.09.010

Marder, E., & Thirumalai, V. (2002). Cellular, synaptic and network effects of neuromodulation. *Neural Networks*, *15*, 479–493.

Mata, R., Wilke, A., & Czienskowski, U. (2013). Foraging across the life span: is there a reduction in exploration with aging? *Frontiers in Neuroscience*, *7*, 53. https://doi.org/10.3389/fnins.2013.00053

MATLAB Release 2014b. Natick, MA: The MathWorks, Inc.

Mattay, V. S., Goldberg, T. E., Fera, F., Hariri, A. R., Tessitore, A., Egan, M. F., . . . Weinberger, D. R. (2003). Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 6186–6191. https://doi.org/10.1073/pnas.0931309100

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., . . . Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin*, *54*, 287–298.

McCabe, C., Huber, A., Harmer, C. J., & Cowen, P. J. (2011). The D2 antagonist sulpiride modulates the neural processing of both rewarding and aversive stimuli in healthy volunteers. *Psychopharmacology*, *217*, 271–278. https://doi.org/10.1007/s00213-011-2278-4

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*, 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

McClure, S. M., Gilzenrat, M. S., & Cohen, J. D. (2006). An exploration-exploitation model based on norepinepherine and dopamine activity. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems, vol. 18* (pp. 867–974). Cambridge, MA: MIT Press.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507. https://doi.org/10.1126/science.1100907

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). New York, NY: Academic Press.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience*, *15*, 394–408. https://doi.org/10.1162/089892903321593117

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., . . . Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215. https://doi.org/10.1037/dec0000033

Mehta, M. A., Hinton, E. C., Montgomery, A. J., Bantick, R. A., & Grasby, P. M. (2005). Sulpiride and mnemonic function: effects of a dopamine D2 receptor antagonist on working memory, emotional memory and long-term memory in healthy volunteers. *Journal of Psychopharmacology*, *19*, 29–38. https://doi.org/10.1177/0269881105048889

Mehta, M. A., Manes, F. F., Magnolfi, G., Sahakian, B. J., & Robbins, T. W. (2004). Impaired set-shifting and dissociable effects on tests of spatial working memory following the dopamine D2 receptor antagonist sulpiride in human volunteers. *Psychopharmacology*, *176*, 331–342. https://doi.org/10.1007/s00213-004-1899-2

Mehta, M. A., McGowan, S. W., Lawrence, A. D., Aitken, M. R. F., Montgomery, A. J., & Grasby, P. M. (2003). Systemic sulpiride modulates striatal blood flow: relationships to spatial working memory and planning. *NeuroImage*, *20*, 1982–1994.

Mehta, M. A., Owen, A. M., Sahakian, B. J., Mavaddat, N., Pickard, J. D., & Robbins, T. W. (2000). Methylphenidate Enhances Working Memory by Modulating Discrete Frontal and Parietal Lobe Regions in the Human Brain. *The Journal of Neuroscience*, *20*, RC65-RC65. https://doi.org/10.1523/JNEUROSCI.20-06-j0004.2000

Mehta, M. A., Sahakian, B. J., McKenna, P. J., & Robbins, T. W. (1999). Systemic sulpiride in young adult volunteers simulates the profile of cognitive deficits in Parkinson's disease. *Psychopharmacology*, *146*, 162–174.

Meiser, J., Weindl, D., & Hiller, K. (2013). Complexity of dopamine metabolism. *Cell Communication and Signaling*, *11*, 34. https://doi.org/10.1186/1478-811X-11-34

Melamed, E., Hefti, F., & Wurtman, R. J. (1980). Nonaminergic striatal neurons convert exogenous L-Dopa to dopamine in parkinsonism. *Annals of Neurology*, *8*, 558–563. https://doi.org/10.1002/ana.410080603

Meller, E., Bohmaker, K., Namba, Y., Friedhoff, A. J., & Goldstein, M. (1987). Relationship between receptor occupancy and response at striatal dopamine autoreceptors. *Molecular Pharmacology*, *31*, 592–598.

Menon, V. (2015). Salience Network. In *Brain Mapping* (pp. 597–611). Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00052-X

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure & Function*, *214*, 655–667. https://doi.org/10.1007/s00429-010-0262-0

Meyer-Lindenberg, A., Kohn, P. D., Kolachana, B., Kippenhan, S., McInerney-Leo, A., Nussbaum, R., . . . Berman, K. F. (2005). Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nature Neuroscience*, *8*, 594–596. https://doi.org/10.1038/nn1438

Miller, D. W., & Abercrombie, E. D. (1999). Role of High-Affinity Dopamine Uptake and Impulse Activity in the Appearance of Extracellular Dopamine in Striatum After Administration of Exogenous L-DOPA. Studies in Intact and 6-Hydroxydopamine-Treated Rats. *Journal of Neurochemistry*, *72*, 1516–1522. https://doi.org/10.1046/j.1471-4159.1999.721516.x

Minzenberg, M. J., Xu, K., Mitropoulou, V., Harvey, P. D., Finch, T., Flory, J. D., . . . Siever, L. J. (2006). Catechol-O-methyltransferase Val158Met genotype variation is associated with prefrontal-dependent task performance in schizotypal personality disorder patients and comparison groups. *Psychiatric Genetics*, *16*, 117–124. https://doi.org/10.1097/01.ypg.0000199448.00163.e6

Missale, C., Nash, S. R., Robinson, S. W., Jaber, M., & Caron, M. G. (1998). Dopamine receptors: from structure to function. *Physiological Reviews*, *78*, 189–225. https://doi.org/10.1152/physrev.1998.78.1.189

Moghaddam, B., & Bunney, B. S. (1990). Acute Effects of Typical and Atypical Antipsychotic Drugs on the Release of Dopamine from Prefrontal Cortex, Nucleus Accumbens, and Striatum of the Rat: An In Vivo Microdialysis Study. *Journal of Neurochemistry*, *54*, 1755–1760. https://doi.org/10.1111/j.1471-4159.1990.tb01230.x

Mohr, C., Sándor, P. S., Landis, T., Fathi, M., & Brugger, P. (2005). Blinking and schizotypal thinking. *Journal of Psychopharmacology*, *19*, 513–520. https://doi.org/10.1177/0269881105056538

Molina-Castillo, F.-J., Jimenez-Jimenez, D., & Munuera-Aleman, J.-L. (2011). Product competence exploitation and exploration strategies: The impact on new product performance through quality and innovativeness. *Industrial Marketing Management*, *40*, 1172–1182. https://doi.org/10.1016/j.indmarman.2010.12.017

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Morris, L. S., Baek, K., Kundu, P., Harrison, N. A., Frank, M. J., & Voon, V. (2016). Biases in the Explore-Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology*, *41*, 940–948. https://doi.org/10.1038/npp.2015.208

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. *The Journal of Neuroscience*, *28*, 12294–12304. https://doi.org/10.1523/JNEUROSCI.3116-08.2008

Müller, J., Dreisbach, G., Brocke, B., Lesch, K.-P., Strobel, A., & Goschke, T. (2007). Dopamine and cognitive control: the influence of spontaneous eyeblink rate, DRD4 exon III polymorphism and gender on flexibility in set-shifting. *Brain Research*, *1131*, 155–162. https://doi.org/10.1016/j.brainres.2006.11.002

Müller, T. (2007). The Role of Levodopa in the Chronic Neurodegenerative Disorder – Parkinson's Disease. In *Oxidative Stress and Neurodegenerative Disorders* (pp. 237–246). Elsevier. https://doi.org/10.1016/B978-044452809-4/50151-4

Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of regressors in FMRI models. *PloS One*, *10*, e0126255. https://doi.org/10.1371/journal.pone.0126255

Mura, A., Jackson, D., Manley, M. S., Young, S. J., & Groves, P. M. (1995). Aromatic L-amino acid decarboxylase immunoreactive cells in the rat striatum: a possible site for the conversion of exogenous L-DOPA to dopamine. *Brain Research*, *704*, 51–60.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, *48*, 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Myöhänen, T. T., & Männistö, P. T. (2010). Distribution and functions of catechol-O-methyltransferase proteins: do recent findings change the picture? *International Review of Neurobiology*, *95*, 29–47. https://doi.org/10.1016/B978-0-12-381326-8.00003-X

Myöhänen, T. T., Schendzielorz, N., & Männistö, P. T. (2010). Distribution of catechol-O-methyltransferase (COMT) proteins and enzymatic activities in wild-type and soluble COMT deficient mice. *Journal of Neurochemistry*, *113*, 1632–1643. https://doi.org/10.1111/j.1471-4159.2010.06723.x

Nadim, F., & Bucher, D. (2014). Neuromodulation of Neurons and Synapses. *Current Opinion in Neurobiology*, *0*, 48–56. https://doi.org/10.1016/j.conb.2014.05.003

Nagao, K., Ding, C., Ganga, G., Alty, J. E., Clissold, B. G., McColl, C. D., . . . Kempster, P. A. (2018). Inferring the long duration response to levodopa in Parkinson's disease. *Parkinsonism & Related Disorders.* Advance online publication. https://doi.org/10.1016/j.parkreldis.2018.09.002

Naqvi, N. H., & Bechara, A. (2009). The hidden island of addiction: the insula. *Trends in Neurosciences*, *32*, 56–67. https://doi.org/10.1016/j.tins.2008.09.009

Narita, M., Matsushima, Y., Niikura, K., Narita, M., Takagi, S., Nakahara, K., . . . Suzuki, T. (2010). Implication of dopaminergic projection from the ventral tegmental area to the anterior cingulate cortex in μ-opioid-induced place preference. *Addiction Biology*, *15*, 434–447. https://doi.org/10.1111/j.1369-1600.2010.00249.x

Naudé, J., Tolu, S., Dongelmans, M., Torquet, N., Valverde, S., Rodriguez, G., . . . Faure, P. (2016). Nicotinic receptors in the ventral tegmental area promote uncertainty-seeking. *Nature Neuroscience*, *19*, 471–478. https://doi.org/10.1038/nn.4223

Navarro, D. J., Newell, B. R., & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore-exploit dilemma in static and dynamic environments. *Cognitive Psychology*, *85*, 43–77. https://doi.org/10.1016/j.cogpsych.2016.01.001

Nestler, E. J. (2005). Is there a common molecular pathway for addiction? *Nature Neuroscience*, *8*, 1445–1449. https://doi.org/10.1038/nn1578

Neve, K. A. (Ed.). (2010). *The receptors*. *The Dopamine Receptors* (2nd ed.). New York, NY: Humana Press.

Neve, K. A., & Neve, R. L. (Eds.). (1997). *The receptors*. *The Dopamine Receptors*. Totowa, N.J.: Humana Press.

Neve, K. A., Seamans, J. K., & Trantham-Davidson, H. (2004). Dopamine Receptor Signaling. *Journal of Receptors and Signal Transduction*, *24*, 165–205. https://doi.org/10.1081/RRS-200029981

Nichols, D. E. (2010). Dopamine Receptor Subtype-Selective Drugs: D1-Like Receptors. In K. A. Neve (Ed.), *The receptors. The Dopamine Receptors* (2nd ed., pp. 75–99). New York, NY: Humana Press.

Nichols, T. E., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*, 419–446. https://doi.org/10.1191/0962280203sm341ra

Nichols, T. E. (2012). Multiple testing corrections, nonparametric methods, and random field theory. *NeuroImage*, *62*, 811–815. https://doi.org/10.1016/j.neuroimage.2012.04.014

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience*, *12*, 241–268. https://doi.org/10.3758/s13415-011-0083-5

Nieratschker, V., Kiefer, C., Giel, K., Krüger, R., & Plewnia, C. (2015). The COMT Val/Met polymorphism modulates effects of tDCS on response inhibition. *Brain Stimulation*, *8*, 283–288. https://doi.org/10.1016/j.brs.2014.11.009

Nissinen, E., & Männistö, P. T. (2010). Biochemistry and pharmacology of catechol-O-methyltransferase inhibitors. *International Review of Neurobiology*, *95*, 73–118. https://doi.org/10.1016/B978-0-12-381326-8.00005-3

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*, 507–520. https://doi.org/10.1007/s00213-006-0502-4

Nord, M. (2017). *Levodopa pharmacokinetics – from stomach to brain: A study on patients with Parkinson's disease*. Linköping: Linköping University Electronic Press.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). *McGraw-Hill series in psychology*. New York, NY: McGraw-Hill.

Nutt, J. G., Carter, J. H., & Woodward, W. R. (1995). Long-duration response to levodopa. *Neurology*, *45*, 1613–1616. https://doi.org/10.1212/WNL.45.8.1613

Nutt, J. G., Woodward, W. R., Hammerstad, J. P., Carter, J. H., & Anderson, J. L. (1984). The "on-off" phenomenon in Parkinson's disease. Relation to levodopa absorption and transport. *The New England Journal of Medicine*, *310*, 483–488. https://doi.org/10.1056/NEJM198402233100802

Nyholm, D., Lewander, T., Gomes-Trolin, C., Bäckström, T., Panagiotidis, G., Ehrnebo, M., . . . Aquilonius, S.-M. (2012). Pharmacokinetics of levodopa/carbidopa microtablets versus levodopa/benserazide and levodopa/carbidopa in healthy volunteers. *Clinical Neuropharmacology*, *35*, 111–117. https://doi.org/10.1097/WNF.0b013e31825645d1

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity – facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045. https://doi.org/10.1016/S0191-8869(99)00251-2

Obeso, J. A., Rodríguez-Oroz, M. C., Benitez-Temino, B., Blesa, F. J., Guridi, J., Marin, C., & Rodriguez, M. (2008). Functional organization of the basal ganglia: therapeutic implications for Parkinson's disease. *Movement Disorders*, *23 Suppl 3*, S548-59. https://doi.org/10.1002/mds.22062

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–249. https://doi.org/10.1016/j.tics.2005.03.010

O'Doherty, J. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*, 769–776. https://doi.org/10.1016/j.conb.2004.10.016

O'Doherty, J. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, *1239*, 118–129. https://doi.org/10.1111/j.1749-6632.2011.06290.x

O'Doherty, J., Dayan, P., Friston, K. J., Critchley, H., & Dolan, R. J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron*, *38*, 329–337. https://doi.org/10.1016/S0896-6273(03)00169-7

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. J., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*, 452–454. https://doi.org/10.1126/science.1094285

O'Doherty, J., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53. https://doi.org/10.1196/annals.1390.022

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5951–5955. https://doi.org/10.1073/pnas.89.13.5951

Ohara, P. T., Granato, A., Moallem, T. M., Wang, B.-R., Tillet, Y., & Jasmin, L. (2003). Dopaminergic input to GABAergic neurons in the rostral agranular insular cortex of the rat. *Journal of Neurocytology*, *32*, 131–141. https://doi.org/10.1023/B:NEUR.0000005598.09647.7f

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Oldham, W. M., & Hamm, H. E. (2008). Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Reviews. Molecular Cell Biology*, *9*, 60–71. https://doi.org/10.1038/nrm2299

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*, 604–609. https://doi.org/10.1093/scan/nss055

O'Sullivan, S. S., Evans, A. H., & Lees, A. J. (2009). Dopamine dysregulation syndrome: an overview of its epidemiology, mechanisms and management. *CNS Drugs*, *23*, 157–170. https://doi.org/10.2165/00023210-200923020-00005

Ovallath, S., & Sulthana, B. (2017). Levodopa: History and Therapeutic Applications. *Annals of Indian Academy of Neurology*, *20*, 185–189. https://doi.org/10.4103/aian.AIAN_241_17

Owesson-White, C. A., Roitman, M. F., Sombers, L. A., Belle, A. M., Keithley, R. B., Peele, J. L., . . . Wightman, R. M. (2012). Sources contributing to the average extracellular concentration of dopamine in the nucleus accumbens. *Journal of Neurochemistry*, *121*, 252–262. https://doi.org/10.1111/j.1471-4159.2012.07677.x

Pajkossy, P., Szőllősi, Á., Demeter, G., & Racsmány, M. (2017). Tonic noradrenergic activity modulates explorative behavior and attentional set shifting: Evidence from pupillometry and gaze pattern analysis. *Psychophysiology*, *54*, 1839–1854. https://doi.org/10.1111/psyp.12964

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive Human Brain Stimulation in Cognitive Neuroscience: A Primer. *Neuron*, *87*, 932–945. https://doi.org/10.1016/j.neuron.2015.07.032

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Computational Biology*, *7*, e1001048. https://doi.org/10.1371/journal.pcbi.1001048

Payzan-LeNestour, E., & Bossaerts, P. (2012). Do not Bet on the Unknown Versus Try to Find Out More: Estimation Uncertainty and "Unexpected Uncertainty" Both Modulate Exploration. *Frontiers in Neuroscience*, *6*, 150. https://doi.org/10.3389/fnins.2012.00150

Pearson, J. M., Hayden, B. Y., Raghavachari, S., & Platt, M. L. (2009). Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Current Biology*, *19*, 1532–1537. https://doi.org/10.1016/j.cub.2009.07.048

Pehek, E. A. (1999). Comparison of effects of haloperidol administration on amphetamine-stimulated dopamine release in the rat medial prefrontal cortex and dorsal striatum. *The Journal of Pharmacology and Experimental Therapeutics*, *289*, 14–23.

Pelchat, M. L., Johnson, A., Chan, R., Valdez, J., & Ragland, J. D. (2004). Images of desire: food-craving activation during fMRI. *NeuroImage*, *23*, 1486–1493. https://doi.org/10.1016/j.neuroimage.2004.08.023

Penny, W. D., & Holmes, A. J. (2006). Random Effects Analysis. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping.* San Diego: Academic Press.

Peretti, C. S., Danion, J. M., Kauffmann-Muller, F., Grangé, D., Patat, A., & Rosenzweig, P. (1997). Effects of haloperidol and amisulpride on motor and cognitive skill learning in healthy volunteers. *Psychopharmacology*, *131*, 329–338.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*, 1042–1045. https://doi.org/10.1038/nature05051

Peters, J., & Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, *213*, 135–141. https://doi.org/10.1016/j.bbr.2010.04.031

Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends in Cognitive Sciences*, *15*, 227–239. https://doi.org/10.1016/j.tics.2011.03.002

Pezze, M. A., Dalley, J. W., & Robbins, T. W. (2009). Remediation of attentional dysfunction in rats with lesions of the medial prefrontal cortex by intra-accumbens administration of the dopamine D(2/3) receptor antagonist sulpiride. *Psychopharmacology*, *202*, 307–313. https://doi.org/10.1007/s00213-008-1384-4

Phillips, P. E. M., Robinson, D. L., Stuber, G. D., Carelli, R. M., & Wightman, R. M. (2002). Real-Time Measurements of Phasic Changes in Extracellular Dopamine Concentration in Freely Moving Rats by Fast-Scan Cyclic Voltammetry. In J. Q. Wang (Ed.), *Drugs of Abuse* (pp. 443–464). New Jersey: Humana Press. https://doi.org/10.1385/1-59259-358-5:443

Pierce, K. L., Premont, R. T., & Lefkowitz, R. J. (2002). Seven-transmembrane receptors. *Nature Reviews. Molecular Cell Biology*, *3*, 639–650. https://doi.org/10.1038/nrm908

Pine, A., Shiner, T., Seymour, B., & Dolan, R. J. (2010). Dopamine, time, and impulsivity in humans. *The Journal of Neuroscience*, *30*, 8888–8896. https://doi.org/10.1523/JNEUROSCI.6028-09.2010

Pisani, L., Catto, M., Leonetti, F., Nicolotti, O., Stefanachi, A., Campagna, F., & Carotti, A. (2011). Targeting Monoamine Oxidases with Multipotent Ligands: An Emerging Strategy in the Search of New Drugs Against Neurodegenerative Diseases. *Current Medicinal Chemistry*, *18*, 4568–4587. https://doi.org/10.2174/092986711797379302

Pizzagalli, D. A., Evins, A. E., Schetter, E. C., Frank, M. J., Pajtas, P. E., Santesso, D. L., & Culhane, M. (2008). Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology*, *196*, 221–232. https://doi.org/10.1007/s00213-007-0957-y

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*, 233–238. https://doi.org/10.1038/22268

Pleger, B., Ruff, C. C., Blankenburg, F., Klöppel, S., Driver, J., & Dolan, R. J. (2009). Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biology*, *7*, e1000164. https://doi.org/10.1371/journal.pbio.1000164

Plenz, D., & Kital, S. T. (1999). A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. *Nature*, *400*, 677–682. https://doi.org/10.1038/23281

Plewnia, C., Zwissler, B., Längst, I., Maurer, B., Giel, K., & Krüger, R. (2013). Effects of transcranial direct current stimulation (tDCS) on executive functions: influence of COMT Val/Met polymorphism. *Cortex*, *49*, 1801–1807. https://doi.org/10.1016/j.cortex.2012.11.002

Poewe, W., Antonini, A., Zijlmans, J. C. M., Burkhard, P. R., & Vingerhoets, F. (2010). Levodopa in the treatment of Parkinson's disease: an old drug still going strong. *Clinical Interventions in Aging*, 229–239. https://doi.org/10.2147/CIA.S6456

Pohjalainen, T., Rinne, J. O., Någren, K., Lehikoinen, P., Anttila, K., Syvälahti, E. K., & Hietala, J. (1998). The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Molecular Psychiatry*, *3*, 256–260.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge: Cambridge University Press.

Pomara, C., Cassano, T., D'Errico, S., Bello, S., Romano, A. D., Riezzo, I., & Serviddio, G. (2012). Data Available on the Extent of Cocaine Use and Dependence: Biochemistry, Pharmacologic Effects and Global Burden of Disease of Cocaine Abusers. *Current Medicinal Chemistry*, *19*, 5647–5657. https://doi.org/10.2174/092986712803988811

Poustchi-Amin, M., Mirowitz, S. A., Brown, J. J., McKinstry, R. C., & Li, T. (2001). Principles and applications of echo-planar imaging: a review for the general radiologist. *Radiographics*, *21*, 767–779. https://doi.org/10.1148/radiographics.21.3.g01ma23767

Prante, O., Dörfler, M., & Gmeiner, P. (2010). Dopamine Receptor Subtype-Selective Drugs: D2-Like Receptors. In K. A. Neve (Ed.), *The receptors. The Dopamine Receptors* (2nd ed., pp. 101–135). New York, NY: Humana Press.

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, *51*, 381–390. https://doi.org/10.1016/j.neuron.2006.06.024

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience*, *28*, 2745–2752. https://doi.org/10.1523/JNEUROSCI.4286-07.2008

Price, A., Filoteo, J. V., & Maddox, W. T. (2009). Rule Based Category Learning in Patients with Parkinson's Disease. *Neuropsychologia*, *47*, 1213–1226. https://doi.org/10.1016/j.neuropsychologia.2009.01.031

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x

Pryor, K. O., & Storer, K. P. (2013). Drugs for Neuropsychiatric Disorders. In *Pharmacology and Physiology for Anesthesia* (pp. 180–207). Elsevier. https://doi.org/10.1016/B978-1-4377-1679-5.00011-9

Pycock, C. J., Kerwin, R. W., & Carter, C. J. (1980). Effect of lesion of cortical dopamine terminals on subcortical dopamine receptors in rats. *Nature*, *286*, 74–77. https://doi.org/10.1038/286074a0

Pyke, G. H. (2018). Optimal Foraging Theory: An Introduction. In B. D. Roitberg (Ed.), *Reference module in life sciences.* Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.01156-0

Quiroga-Varela, A., Walters, J. R., Brazhnik, E., Marin, C., & Obeso, J. A. (2013). What basal ganglia changes underlie the parkinsonian state? The significance of neuronal oscillatory activity. *Neurobiology of Disease*, *58*, 242–248. https://doi.org/10.1016/j.nbd.2013.05.010

R Core Team. (2017). R: A Language and Environment for Statistical Computing, version 3.4.3. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Racey, D., Young, M. E., Garlick, D., Pham, J. N.-M., & Blaisdell, A. P. (2011). Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules. *Learning & Behavior*, *39*, 245–258. https://doi.org/10.3758/s13420-011-0025-7

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 676–682. https://doi.org/10.1073/pnas.98.2.676

Raja Beharelle, Anjali, Polanía, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration-Exploitation Trade-Offs. *The Journal of Neuroscience*, *35*, 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015

Ramayya, A. G., Misra, A., Baltuch, G. H., & Kahana, M. J. (2014). Microstimulation of the human substantia nigra alters reinforcement learning. *The Journal of Neuroscience*, *34*, 6887–6895. https://doi.org/10.1523/JNEUROSCI.5445-13.2014

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews. Neuroscience*, *5*, 184–194. https://doi.org/10.1038/nrn1343

Rathelot, J.-A., Dum, R. P., & Strick, P. L. (2017). Posterior parietal cortex contains a command apparatus for hand movements. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 4255–4260. https://doi.org/10.1073/pnas.1608132114

Ray, W. A., Chung, C. P., Murray, K. T., Hall, K., & Stein, C. M. (2009). Atypical Antipsychotic Drugs and the Risk of Sudden Cardiac Death. *The New England Journal of Medicine*, *360*, 225–235. https://doi.org/10.1056/NEJMoa0806994

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring Working Memory Capacity with Automated Complex Span Tasks. *European Journal of Psychological Assessment*, *28*, 164–171. https://doi.org/10.1027/1015-5759/a000123

Reid, C. R., MacDonald, H., Mann, R. P., Marshall, J. A. R., Latty, T., & Garnier, S. (2016). Decision-making without a brain: how an amoeboid organism solves the two-armed bandit. *Journal of the Royal Society, Interface*, *13.* https://doi.org/10.1098/rsif.2016.0030

Richerson, G. B., Aston-Jones, G., & Saper, C. B. (2013). The Modulatory Functions of the Brain Stem. In E. Kandel (Ed.), *Principles of neural science* (5th ed., pp. 1038–1055). New York, NY: McGraw-Hill.

Richfield, E. K., Young, A. B., & Penney, J. B. (1989). Comparative distributions of dopamine D-1 and D-2 receptors in the cerebral cortex of rats, cats, and monkeys. *The Journal of Comparative Neurology*, *286*, 409–426. https://doi.org/10.1002/cne.902860402

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447. https://doi.org/10.1126/science.1100301

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*, 129–140. https://doi.org/10.1016/j.bandc.2004.09.016

Riederer, P., & Laux, G. (2011). MAO-inhibitors in Parkinson's Disease. *Experimental Neurobiology*, *20*, 1–17. https://doi.org/10.5607/en.2011.20.1.1

Rigoli, F., Chew, B., Dayan, P., & Dolan, R. J. (2016). The Dopaminergic Midbrain Mediates an Effect of Average Reward on Pavlovian Vigor. *Journal of Cognitive Neuroscience*, *28*, 1303–1317. https://doi.org/10.1162/jocn_a_00972

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, *58*, 527–535.

Robert, C. P., & Casella, G. (2005). *Monte Carlo statistical methods (2nd ed.)*. *Springer texts in statistics*. New York: Springer.

Roberts, A. C., de Salvia, M. A., Wilkinson, L. S., Collins, P., Muir, J. L., Everitt, B. J., & Robbins, T. W. (1994). 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin Card Sort Test: possible interactions with subcortical dopamine. *The Journal of Neuroscience*, *14*, 2531–2544.

Robinson, D. L. (2003). Detecting Subsecond Dopamine Release with Fast-Scan Cyclic Voltammetry in Vivo. *Clinical Chemistry*, *49*, 1763–1773. https://doi.org/10.1373/49.10.1763

Rollema, H., Lu, Y., Schmidt, A. W., Sprouse, J. S., & Zorn, S. H. (2000). 5-HT(1A) receptor activation contributes to ziprasidone-induced dopamine release in the rat prefrontal cortex. *Biological Psychiatry*, *48*, 229–237.

Romanelli, R. J., Williams, J. T., & Neve, K. A. (2010). Dopamine Receptor Signaling: Intracellular Pathways to Behavior. In K. A. Neve (Ed.), *The receptors. The Dopamine Receptors* (2nd ed.). New York, NY: Humana Press.

Rosa, A., Peralta, V., Cuesta, M. J., Zarzuela, A., Serrano, F., Martínez-Larrea, A., & Fañanás, L. (2004). New evidence of association between COMT gene and prefrontal neurocognitive function in healthy individuals from sibling pairs discordant for psychosis. *The American Journal of Psychiatry*, *161*, 1110–1112. https://doi.org/10.1176/appi.ajp.161.6.1110

Roth, R. H. (1984). Cns dopamine autoreceptors: distribution, pharmacology, and function. *Annals of the New York Academy of Sciences*, *430*, 27–53.

Rudorf, S., Preuschoff, K., & Weber, B. (2012). Neural correlates of anticipation risk reflect risk preferences. *The Journal of Neuroscience*, *32*, 16683–16692. https://doi.org/10.1523/JNEUROSCI.4235-11.2012

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*, 389–397. https://doi.org/10.1038/nn2066

Rushworth, M. F. S., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*, 946–955. https://doi.org/10.1016/j.conb.2012.04.011

Ruskin, D. N., Bergstrom, D. A., Kaneoke, Y., Patel, B. N., Twery, M. J., & Walters, J. R. (1999). Multisecond oscillations in firing rate in the basal ganglia: robust modulation by dopamine receptor activation and anesthesia. *Journal of Neurophysiology*, *81*, 2046–2055. https://doi.org/10.1152/jn.1999.81.5.2046

Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *The Journal of Neuroscience*, *29*, 15104–15114. https://doi.org/10.1523/JNEUROSCI.3524-09.2009

Saddoris, M. P., Sugam, J. A., Stuber, G. D., Witten, I. B., Deisseroth, K., & Carelli, R. M. (2015). Mesolimbic dopamine dynamically tracks, and is causally linked to, discrete aspects of value-based decision making. *Biological Psychiatry*, *77*, 903–911. https://doi.org/10.1016/j.biopsych.2014.10.024

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776. https://doi.org/10.1037//0012-1649.27.5.763

Sanberg, P. R. (1980). Haloperidol-induced catalepsy is mediated by postsynaptic dopamine receptors. *Nature*, *284*, 472–473. https://doi.org/10.1038/284472a0

Sawa, A., & Snyder, S. H. (2002). Schizophrenia: diverse approaches to a complex disease. *Science*, *296*, 692–695. https://doi.org/10.1126/science.1070532

Sawaguchi, T. (2001). The effects of dopamine and its antagonists on directional delay-period activity of prefrontal neurons in monkeys during an oculomotor delayed-response task. *Neuroscience Research*, *41*, 115–128.

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science*, *251*, 947–950.

Schacht, J. P. (2016). COMT val158met moderation of dopaminergic drug effects on cognitive function: a critical review. *The Pharmacogenomics Journal*, *16*, 430–438. https://doi.org/10.1038/tpj.2016.43

Scheggia, D., Sannino, S., Scattoni, M. L., & Papaleo, F. (2012). COMT as a Drug Target for Cognitive Functions and Dysfunctions. *CNS & Neurological Disorders – Drug Targets*, *11*, 209–221. https://doi.org/10.2174/187152712800672481

Schinkel, A. H., Wagenaar, E., Mol, C. A., & van Deemter, L. (1996). P-glycoprotein in the blood-brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *Journal of Clinical Investigation*, *97*, 2517–2524. https://doi.org/10.1172/JCI118699

Schmitz, Y., Benoit-Marand, M., Gonon, F., & Sulzer, D. (2003). Presynaptic regulation of dopaminergic neurotransmission. *Journal of Neurochemistry*, *87*, 273–289. https://doi.org/10.1046/j.1471-4159.2003.02050.x

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *The Journal of Neuroscience*, *27*, 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schott, B. H., Frischknecht, R., Debska-Vielhaber, G., John, N., Behnisch, G., Düzel, E., . . . Seidenbecher, C. I. (2010). Membrane-Bound Catechol-O-Methyl Transferase in Cortical Neurons and Glial Cells is Intracellularly Oriented. *Frontiers in Psychiatry*, *1*, 142. https://doi.org/10.3389/fpsyt.2010.00142

Schrantee, A., & Reneman, L. (2014). Pharmacological imaging as a tool to visualise dopaminergic neurotoxicity. *Neuropharmacology*, *84*, 159–169. https://doi.org/10.1016/j.neuropharm.2013.06.029

Schultz, W. (2002). Getting Formal with Dopamine and Reward. *Neuron*, *36*, 241–263. https://doi.org/10.1016/S0896-6273(02)00967-4

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annual Review of Neuroscience*, *30*, 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722

Schultz, W. (2010). Dopamine signals for reward value and risk: basic and recent data. *Behavioral and Brain Functions*, *6*, 24. https://doi.org/10.1186/1744-9081-6-24

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*, 23–32.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience*, *13*, 900–913.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*, 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schultz, W., Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C. D., Tobler, P. N., & Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*, 3801–3811. https://doi.org/10.1098/rstb.2008.0152

Schultz, W., Stauffer, W. R., & Lak, A. (2017). The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Current Opinion in Neurobiology*, *43*, 139–148. https://doi.org/10.1016/j.conb.2017.03.013

Schwarz, A., Gozzi, A., Reese, T., Bertani, S., Crestan, V., Hagan, J., . . . Bifone, A. (2004). Selective dopamine D(3) receptor antagonist SB-277011-A potentiates phMRI response to acute amphetamine challenge in the rat brain. *Synapse*, *54*, 1–10. https://doi.org/10.1002/syn.20055

Schweimer, J., & Hauber, W. (2006). Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learning & Memory*, *13*, 777–782. https://doi.org/10.1101/lm.409306

Schwerdt, H. N., Shimazu, H., Amemori, K.-I., Amemori, S., Tierney, P. L., Gibson, D. J., . . . Graybiel, A. M. (2017). Long-term dopamine neurochemical monitoring in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 13260–13265. https://doi.org/10.1073/pnas.1713756114

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, *74*, 1–58. https://doi.org/10.1016/j.pneurobio.2004.05.006

Sesack, S. R., & Grace, A. A. (2010). Cortico-Basal Ganglia reward network: microcircuitry. *Neuropsychopharmacology*, *35*, 27–47. https://doi.org/10.1038/npp.2009.93

Sesack, S. R., Hawrylak, V. A., Matus, C., Guido, M. A., & Levey, A. I. (1998). Dopamine Axon Varicosities in the Prelimbic Division of the Rat Prefrontal Cortex Exhibit Sparse Immunoreactivity for the Dopamine Transporter. *The Journal of Neuroscience*, *18*, 2697–2708. https://doi.org/10.1523/JNEUROSCI.18-07-02697.1998

Sescousse, G., Ligneul, R., van Holst, R. J., Janssen, L. K., de Boer, F., Janssen, M., . . . Cools, R. (2018). Spontaneous eye blink rate and dopamine synthesis capacity: Preliminary evidence for an absence of positive correlation. *European Journal of Neuroscience*, *47*, 1081–1086. https://doi.org/10.1111/ejn.13895

Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron*, *90*, 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Shen, H.-W., Hagino, Y., Kobayashi, H., Shinohara-Tanaka, K., Ikeda, K., Yamamoto, H., . . . Sora, I. (2004). Regional differences in extracellular dopamine and serotonin assessed by in vivo microdialysis in mice lacking dopamine and/or serotonin transporters. *Neuropsychopharmacology*, *29*, 1790–1799. https://doi.org/10.1038/sj.npp.1300476

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*, 1286–1291. https://doi.org/10.1038/nn.4384

Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, *17*, 1249–1254. https://doi.org/10.1038/nn.3771

Sherman, E., & Wilson, R. C. (2016). Spontaneous Blink Rate Correlates With Financial Risk Taking. *BioRxiv.* Advance online publication. https://doi.org/10.1101/046821

Shi, W.-X., Smith, P. L., Pun, C.-L., Millet, B., & Bunney, B. S. (1997). D1–D2 Interaction in Feedback Control of Midbrain Dopamine Neurons. *The Journal of Neuroscience*, *17*, 7988–7994. https://doi.org/10.1523/JNEUROSCI.17-20-07988.1997

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Simon, N. W., Montgomery, K. S., Beas, B. S., Mitchell, M. R., LaSarge, C. L., Mendez, I. A., . . . Setlow, B. (2011). Dopaminergic modulation of risky decision-making. *The Journal of Neuroscience*, *31*, 17460–17470. https://doi.org/10.1523/JNEUROSCI.3772-11.2011

Simpson, E. H., & Kellendonk, C. (2017). Insights About Striatal Circuit Function and Schizophrenia From a Mouse Model of Dopamine D2 Receptor Upregulation. *Biological Psychiatry*, *81*, 21–30. https://doi.org/10.1016/j.biopsych.2016.07.004

Simpson, E. H., Kellendonk, C., & Kandel, E. (2010). A possible role for the striatum in the pathogenesis of the cognitive symptoms of schizophrenia. *Neuron*, *65*, 585–596. https://doi.org/10.1016/j.neuron.2010.02.014

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*, 334–340. https://doi.org/10.1016/j.tics.2009.05.001

Slagter, H. A., Georgopoulou, K., & Frank, M. J. (2015). Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia*, *71*, 126–132. https://doi.org/10.1016/j.neuropsychologia.2015.03.028

Slagter, H. A., Tomer, R., Christian, B. T., Fox, A. S., Colzato, L. S., King, C. R., . . . Davidson, R. J. (2012). PET evidence for a role for striatal dopamine in the attentional blink: functional implications. *Journal of Cognitive Neuroscience*, *24*, 1932–1940. https://doi.org/10.1162/jocn_a_00255

Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics*, *9*, 477–485. https://doi.org/10.1038/nrg2361

Sławek, J., Derejko, M., & Lass, P. (2005). Factors affecting the quality of life of patients with idiopathic Parkinson's disease – a cross-sectional study in an outpatient clinic attendees. *Parkinsonism & Related Disorders*, *11*, 465–468. https://doi.org/10.1016/j.parkreldis.2005.04.006

Smith, A. D., Olson, R. J., & Justice, J. B. (1992). Quantitative microdialysis of dopamine in the striatum: effect of circadian variation. *Journal of Neuroscience Methods*, *44*, 33–41. https://doi.org/10.1016/0165-0270(92)90111-P

Smith, Y., & Villalba, R. (2008). Striatal and extrastriatal dopamine in the basal ganglia: an overview of its anatomical organization in normal and Parkinsonian brains. *Movement Disorders*, *23 Suppl 3*, S534-47. https://doi.org/10.1002/mds.22027

Soares, J. M., Magalhães, R., Moreira, P. S., Sousa, A., Ganz, E., Sampaio, A., . . . Sousa, N. (2016). A Hitchhiker's Guide to Functional Magnetic Resonance Imaging. *Frontiers in Neuroscience*, *10*, 515. https://doi.org/10.3389/fnins.2016.00515

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology. General*, *146*, 155–164. https://doi.org/10.1037/xge0000250

Sotnikova, T. D., Beaulieu, J.-M., Gainetdinov, R. R., & Caron, M. G. (2006). Molecular biology, pharmacology and functional role of the plasma membrane dopamine transporter. *CNS & Neurological Disorders – Drug Targets*, *5*, 45–56.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*, 351–367. https://doi.org/10.1111/tops.12145

Sprawls, P. (2000). *Magnetic resonance imaging: Principles, methods and techniques*. Madison, Wiscon.: Medical physics Publishing.

St Onge, J. R., Abhari, H., & Floresco, S. B. (2011). Dissociable contributions by prefrontal D1 and D2 receptors to risk-based decision making. *The Journal of Neuroscience*, *31*, 8625–8633. https://doi.org/10.1523/JNEUROSCI.1020-11.2011

St Onge, J. R., & Floresco, S. B. (2010). Prefrontal cortical contribution to risk-based decision making. *Cerebral Cortex*, *20*, 1816–1828. https://doi.org/10.1093/cercor/bhp250

Stahl, S. M., & Felker, A. (2008). Monoamine oxidase inhibitors: a modern guide to an unrequited class of antidepressants. *CNS Spectrums*, *13*, 855–870.

Stamford, J. A., Kruk, Z. L., & Millar, J. (1991). Differential effects of dopamine agonists upon stimulated limbic and striatal dopamine release: in vivo voltammetric data. *British Journal of Pharmacology*, *102*, 45–50.

Stan Development Team. (2017a). RStan: the R interface to Stan. R package version 2.17.2. Retrieved from http://mc-stan.org

Stan Development Team. (2017b). Stan Modeling Language Users Guide and Reference Manual, version 2.17.0. Retrieved from http://mc-stan.org

Starke, K., Göthert, M., & Kilbinger, H. (1989). Modulation of neurotransmitter release by presynaptic autoreceptors. *Physiological Reviews*, *69*, 864–989. https://doi.org/10.1152/physrev.1989.69.3.864

Steeves, T. D. L., Miyasaki, J., Zurowski, M., Lang, A. E., Pellecchia, G., van Eimeren, T., . . . Strafella, A. P. (2009). Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a 11C raclopride PET study. *Brain*, *132*, 1376–1385. https://doi.org/10.1093/brain/awp054

Steinberg, E. E., & Janak, P. H. (2013). Establishing causality for dopamine in neural function and behavior with optogenetics. *Brain Research*, *1511*, 46–64. https://doi.org/10.1016/j.brainres.2012.09.036

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*, 966–973. https://doi.org/10.1038/nn.3413

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*, 446–459. https://doi.org/10.1002/wcs.58

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current Opinion in Neurobiology*, *25*, 85–92. https://doi.org/10.1016/j.conb.2013.12.007

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. *Monographs in behavior and ecology*. Princeton, NJ: Princeton University Press.

Stępnicki, P., Kondej, M., & Kaczor, A. A. (2018). Current Concepts and Treatments of Schizophrenia. *Molecules*, *23.* https://doi.org/10.3390/molecules23082087

Stevens, J. R. (1978). Eye blink and schizophrenia: psychosis or tardive dyskinesia? *The American Journal of Psychiatry*, *135*, 223–226. https://doi.org/10.1176/ajp.135.2.223

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*, 168–179. https://doi.org/10.1016/j.jmp.2008.11.002

Stipanovich, A., Valjent, E., Matamales, M., Nishi, A., Ahn, J.-H., Maroteaux, M., . . . Girault, J.-A. (2008). A phosphatase cascade by which rewarding stimuli control nucleosomal response. *Nature*, *453*, 879–884. https://doi.org/10.1038/nature06994

Strange, P. G., & Neve, K. (2013). *Dopamine Receptors*. Tocris Cookson, Ltd. Retrieved from https://resources.tocris.com/pdfs/literature/reviews/dopamine-receptors-review-2018.pdf

Strauss, G. P., Frank, M. J., Waltz, J. A., Kasanova, Z., Herbener, E. S., & Gold, J. M. (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. *Biological Psychiatry*, *69*, 424–431. https://doi.org/10.1016/j.biopsych.2010.10.015

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*, 1782–1787. https://doi.org/10.1126/science.1094765

Sulzer, D. (2011). How addictive drugs disrupt presynaptic dopamine neurotransmission. *Neuron*, *69*, 628–649. https://doi.org/10.1016/j.neuron.2011.02.010

Sulzer, D., Cragg, S. J., & Rice, M. E. (2016). Striatal dopamine neurotransmission: regulation of release and uptake. *Basal Ganglia*, *6*, 123–148. https://doi.org/10.1016/j.baga.2016.02.001

Sutton, R. S. (1990). Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming. In *Machine Learning Proceedings 1990: Proceedings of the Seventh International Conference on Machine Learning, University of Texas, Austin, Texas, June 21-23, 1990* (pp. 216–224). s.l.: Elsevier Reference Monographs. https://doi.org/10.1016/B978-1-55860-141-3.50030-4

Sutton, R. S., & Barto, A. (1998). *Reinforcement learning: An introduction*. *A Bradford book*. Cambridge, MA, London: The MIT Press.

Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction* (2nd ed.). *Adaptive computation and machine learning series*. Cambridge, MA, Lodon: The MIT Press.

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*, e2000797. https://doi.org/10.1371/journal.pbio.2000797

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, *7*, 887–893. https://doi.org/10.1038/nn1279

Tedroff, J., Pedersen, M., Aquilonius, S.-M., Hartvig, P., Jacobsson, G., & Langstrom, B. (1996). Levodopa-induced changes in synaptic dopamine in patients with Parkinson's disease as measured by [11C]raclopride displacement and PET. *Neurology*, *46*, 1430. https://doi.org/10.1212/WNL.46.5.1430

Tharp, I. J., & Pickering, A. D. (2011). Individual differences in cognitive-flexibility: the influence of spontaneous eyeblink rate, trait psychoticism and working memory on attentional set-shifting. *Brain and Cognition*, *75*, 119–125. https://doi.org/10.1016/j.bandc.2010.10.010

Thompson, J., Thomas, N., Singleton, A., Piggott, M., Lloyd, S., Perry, E. K., . . . Court, J. A. (1997). D2 dopamine receptor gene (DRD2) Taq1 A polymorphism: reduced dopamine D2 receptor binding in the human striatum associated with the A1 allele. *Pharmacogenetics*, *7*, 479–484.

Thrun, S. (1992). The Role of Exploration in Learning Control. In *Handbook for Intelligent Control: Neural, Fuzzy and Adaptive Approaches.* Florence, Kentucky: Van Nostrand Reinhold.

Tian, C., Gregersen, P. K., & Seldin, M. F. (2008). Accounting for ancestry: population substructure and genome-wide association studies. *Human Molecular Genetics*, *17*, R143-50. https://doi.org/10.1093/hmg/ddn268

Tokic, M. (2010). Adaptive ε-Greedy Exploration in Reinforcement Learning Based on Value Differences. In R. Dillmann, J. Beyerer, U. D. Hanebeck, & T. Schultz (Eds.), *Lecture Notes in Computer Science. KI 2010: Advances in Artificial Intelligence* (Vol. 6359, pp. 203–210). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16111-7_23

Tokic, M., & Palm, G. (2011). Value-Difference Based Exploration: Adaptive Control between Epsilon-Greedy and Softmax. In J. Bach & S. Edelkamp (Eds.), *Lecture Notes in Computer Science. KI 2011: Advances in Artificial Intelligence* (Vol. 7006, pp. 335–346). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24455-1_33

Torres, G. E., Gainetdinov, R. R., & Caron, M. G. (2003). Plasma membrane monoamine transporters: structure, regulation and function. *Nature Reviews. Neuroscience*, *4*, 13–25. https://doi.org/10.1038/nrn1008

Tost, H., Alam, T., & Meyer-Lindenberg, A. (2010). Dopamine and psychosis: theory, pathomechanisms and intermediate phenotypes. *Neuroscience and Biobehavioral Reviews*, *34*, 689–700. https://doi.org/10.1016/j.neubiorev.2009.06.005

Trantham-Davidson, H., Neely, L. C., Lavin, A., & Seamans, J. K. (2004). Mechanisms underlying differential D1 versus D2 dopamine receptor regulation of inhibition in prefrontal cortex. *The Journal of Neuroscience*, *24*, 10652–10659. https://doi.org/10.1523/JNEUROSCI.3179-04.2004

Tritsch, N. X., & Sabatini, B. L. (2012). Dopaminergic modulation of synaptic transmission in cortex and striatum. *Neuron*, *76*, 33–50. https://doi.org/10.1016/j.neuron.2012.09.023

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., de Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, *324*, 1080–1084. https://doi.org/10.1126/science.1168878

Tsuang, M. (2000). Schizophrenia: genes and environment. *Biological Psychiatry*, *47*, 210–220.

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: encoding ends at the end of the endbrain. *Trends in Cognitive Sciences*, *15*, 169–176. https://doi.org/10.1016/j.tics.2011.02.001

Tunbridge, E. M., Bannerman, D. M., Sharp, T., & Harrison, P. J. (2004). Catechol-o-methyltransferase inhibition improves set-shifting performance and elevates stimulated dopamine release in the rat prefrontal cortex. *The Journal of Neuroscience*, *24*, 5331–5335. https://doi.org/10.1523/JNEUROSCI.1124-04.2004

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*, 62. https://doi.org/10.1038/s42003-018-0073-z

Tversky, A., & Edwards, W. (1966). Information versus reward in binary choices. *Journal of Experimental Psychology*, *71*, 680–683.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews. Neuroscience*, *16*, 55–61. https://doi.org/10.1038/nrn3857

Ueda, Y., Tominaga, A., Kajimura, S., & Nomura, M. (2016). Spontaneous eye blinks during creative task correlate with divergent processing. *Psychological Research*, *80*, 652–659. https://doi.org/10.1007/s00426-015-0665-x

Ugrumov, M. V. (2009). Non-dopaminergic neurons partly expressing dopaminergic phenotype: distribution in the brain, development and functional significance. *Journal of Chemical Neuroanatomy*, *38*, 241–256. https://doi.org/10.1016/j.jchemneu.2009.08.004

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: a latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654. https://doi.org/10.1080/09658210902998047

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: analysis of S&P 500 corporations. *Strategic Management Journal*, *30*, 221–231. https://doi.org/10.1002/smj.738

Van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A Standard Computerized Version of the Reading Span Test in Different Languages. *European Journal of Psychological Assessment*, *24*, 35–42. https://doi.org/10.1027/1015-5759.24.1.35

Van der Post, J., de Waal, P. P., de Kam, M. L., Cohen, A. F., & van Gerven, J. M. A. (2004). No evidence of the usefulness of eye blinking as a marker for central dopaminergic activity. *Journal of Psychopharmacology*, *18*, 109–114. https://doi.org/10.1177/0269881104042832

Van der Schaaf, M. E., van Schouwenburg, M. R., Geurts, D. E. M., Schellekens, A. F. A., Buitelaar, J. K., Verkes, R. J., & Cools, R. (2014). Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. *Cerebral Cortex*, *24*, 633–642. https://doi.org/10.1093/cercor/bhs344

Van Eimeren, T., Ballanger, B., Pellecchia, G., Miyasaki, J. M., Lang, A. E., & Strafella, A. P. (2009). Dopamine agonists diminish value sensitivity of the orbitofrontal cortex: a trigger for pathological gambling in Parkinson's disease? *Neuropsychopharmacology*, *34*, 2758–2766. https://doi.org/10.1038/sj.npp.npp2009124

Van Holstein, M., Froböse, M. I., O'Shea, J., Aarts, E., & Cools, R. (2018). Controlling striatal function via anterior frontal cortex stimulation. *Scientific Reports*, *8*, 3312. https://doi.org/10.1038/s41598-018-21346-5

Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2018). Pupil responses as indicators of value-based decision-making. *BioRxiv.* Advance online publication. https://doi.org/10.1101/302166

Van Veen, V., & Carter, C. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*, 477–482. https://doi.org/10.1016/S0031-9384(02)00930-7

Van Velzen, L. S., Vriend, C., de Wit, S. J., & van den Heuvel, O. A. (2014). Response inhibition and interference control in obsessive-compulsive spectrum disorders. *Frontiers in Human Neuroscience*, *8*, 419. https://doi.org/10.3389/fnhum.2014.00419

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vermorel, J., & Mohri, M. (2005). Multi-armed Bandit Algorithms and Empirical Evaluation. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, . . . L. Torgo (Eds.), *Lecture Notes in Computer Science. Machine Learning: ECML 2005* (Vol. 3720, pp. 437–448). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11564096_42

Volonté, M., Monferini, E., Cerutti, M., Fodritto, F., & Borsini, F. (1997). Bimg 80, a novel potential antipsychotic drug: evidence for multireceptor actions and preferential release of dopamine in prefrontal cortex. *Journal of Neurochemistry*, *69*, 182–190.

Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J., & Hallett, M. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron*, *65*, 135–142. https://doi.org/10.1016/j.neuron.2009.12.027

Voon, V., Reynolds, B., Brezing, C., Gallea, C., Skaljic, M., Ekanayake, V., . . . Hallett, M. (2010). Impulsive choice and response in dopamine agonist-related impulse control behaviors. *Psychopharmacology*, *207*, 645–659. https://doi.org/10.1007/s00213-009-1697-y

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*, 43–48. https://doi.org/10.1038/35083500

Walton, M. E., Kennerley, S. W., Bannerman, D. M., Phillips, P. E. M., & Rushworth, M. F. S. (2006). Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Networks*, *19*, 1302–1314. https://doi.org/10.1016/j.neunet.2006.03.005

Wang, L., Xiong, N., Huang, J., Guo, S., Liu, L., Han, C., . . . Wang, T. (2017). Protein-Restricted Diets for Ameliorating Motor Fluctuations in Parkinson's Disease. *Frontiers in Aging Neuroscience*, *9*, 206. https://doi.org/10.3389/fnagi.2017.00206

Wang, X.-J., & Krystal, J. H. (2014). Computational psychiatry. *Neuron*, *84*, 638–654. https://doi.org/10.1016/j.neuron.2014.10.018

Warren, C. M., Wilson, R. C., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., Cohen, J. D., & Nieuwenhuis, S. (2017). The effect of atomoxetine on random and directed exploration in humans. *PloS One*, *12*, e0176034. https://doi.org/10.1371/journal.pone.0176034

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*, 550–564. https://doi.org/10.3758/BF03195534

Watkinson, S. C., Boddy, L., Burton, K., Darrah, P. R., Eastwood, D., Fricker, M. D., & Tlalka, M. (2005). New approaches to investigating the function of mycelial networks. *Mycologist*, *19*, 11–17. https://doi.org/10.1017/S0269915X05001023

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.

Weinberger, D. R., Egan, M. F., Bertolino, A., Callicott, J. H., Mattay, V. S., Lipska, B. K., . . . Goldberg, T. E. (2001). Prefrontal neurons and the genetics of schizophrenia. *Biological Psychiatry*, *50*, 825–844. https://doi.org/10.1016/S0006-3223(01)01252-5

Weinstein, J. J., Chohan, M. O., Slifstein, M., Kegeles, L. S., Moore, H., & Abi-Dargham, A. (2017). Pathway-Specific Dopamine Abnormalities in Schizophrenia. *Biological Psychiatry*, *81*, 31–42. https://doi.org/10.1016/j.biopsych.2016.03.2104

Weintraub, D. (2008). Dopamine and impulse control disorders in Parkinson's disease. *Annals of Neurology*, *64 Suppl 2*, S93-100. https://doi.org/10.1002/ana.21454

Weintraub, D., Koester, J., Potenza, M. N., Siderowf, A. D., Stacy, M., Voon, V., . . . Lang, A. E. (2010). Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients. *Archives of Neurology*, *67*, 589–595. https://doi.org/10.1001/archneurol.2010.65

Westerink, B. H. (2002). Can antipsychotic drugs be classified by their effects on a particular group of dopamine neurons in the brain? *European Journal of Pharmacology*, *455*, 1–18. https://doi.org/10.1016/S0014-2999(02)02496-2

Westerink, B. H., Kawahara, Y., de Boer, P., Geels, C., de Vries, J. B., Wikström, H. V., . . . Long, S. K. (2001). Antipsychotic drugs classified by their effects on the release of dopamine and noradrenaline in the prefrontal cortex and striatum. *European Journal of Pharmacology*, *412*, 127–138.

Westerink, R. H. S. (2006). Targeting exocytosis: ins and outs of the modulation of quantal dopamine release. *CNS & Neurological Disorders – Drug Targets*, *5*, 57–77.

Westlund, K. N., Denney, R. M., Rose, R. M., & Abell, C. W. (1988). Localization of distinct monoamine oxidase a and monoamine oxidase b cell populations in human brainstem. *Neuroscience*, *25*, 439–456. https://doi.org/10.1016/0306-4522(88)90250-3

Wiegand, A., Nieratschker, V., & Plewnia, C. (2016). Genetic Modulation of Transcranial Direct Current Stimulation Effects on Cognition. *Frontiers in Human Neuroscience*, *10*, 651. https://doi.org/10.3389/fnhum.2016.00651

Wightman, R. M., Amatore, C., Engstrom, R. C., Hale, P. D., Kristensen, E. W., Kuhr, W. G., & May, L. J. (1988). Real-time characterization of dopamine overflow and uptake in the rat striatum. *Neuroscience*, *25*, 513–523.

Wightman, R. M., & Robinson, D. L. (2002). Transient changes in mesolimbic dopamine and their association with 'reward'. *Journal of Neurochemistry*, *82*, 721–735. https://doi.org/10.1046/j.1471-4159.2002.01005.x

Williams, G. V., & Castner, S. A. (2006). Under the curve: critical issues for elucidating D1 receptor function in working memory. *Neuroscience*, *139*, 263–276. https://doi.org/10.1016/j.neuroscience.2005.09.028

Williams, G. V., & Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature*, *376*, 572–575. https://doi.org/10.1038/376572a0

Williams-Gray, C. H., Hampshire, A., Barker, R. A., & Owen, A. M. (2008). Attentional control in Parkinson's disease is dependent on COMT val 158 met genotype. *Brain*, *131*, 397–408. https://doi.org/10.1093/brain/awm313

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology. General*, *143*, 2074–2081. https://doi.org/10.1037/a0038199

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270–273. https://doi.org/10.1126/science.1223252

Wise, R. A. (2005). Forebrain substrates of reward and motivation. *The Journal of Comparative Neurology*, *493*, 115–121. https://doi.org/10.1002/cne.20689

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., . . . Deisseroth, K. (2011). Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*, *72*, 721–733. https://doi.org/10.1016/j.neuron.2011.10.028

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, *58*, 967–973. https://doi.org/10.1016/j.neuron.2008.04.027

Wolf, M. E., & Roth, R. H. (1990). Autoreceptor Regulation of Dopamine Synthesis. *Annals of the New York Academy of Sciences*, *604*, 323–343. https://doi.org/10.1111/j.1749-6632.1990.tb32003.x

Worsley, K. (2006). Random Field Theory. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping.* San Diego: Academic Press.

Worthy, D. A., Pang, B., & Byrne, K. A. (2013). Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Frontiers in Psychology*, *4*, 640. https://doi.org/10.3389/fpsyg.2013.00640

Wu, K., O'Keeffe, D., Politis, M., O'Keeffe, G. C., Robbins, T. W., Bose, S. K., . . . Barker, R. A. (2012). The catechol-O-methyltransferase Val(158)Met polymorphism modulates fronto-cortical dopamine turnover in early Parkinson's disease: a PET study. *Brain*, *135*, 2449–2457. https://doi.org/10.1093/brain/aws157

Wu, Q., Reith, M. E. A., Walker, Q. D., Kuhn, C. M., Carroll, F. I., & Garris, P. A. (2002). Concurrent autoreceptor-mediated control of dopamine release and uptake during neurotransmission: an in vivo voltammetric study. *The Journal of Neuroscience*, *22*, 6272–6281.

Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, *75*, 418–424. https://doi.org/10.1016/j.neuron.2012.03.042

Xu, T.-X., Sotnikova, T. D., Liang, C., Zhang, J., Jung, J. U., Spealman, R. D., . . . Yao, W.-D. (2009). Hyperdopaminergic tone erodes prefrontal long-term potential via a D2 receptor-operated protein phosphatase gate. *The Journal of Neuroscience*, *29*, 14086–14099. https://doi.org/10.1523/JNEUROSCI.0974-09.2009

Xue, G., Lu, Z., Levin, I. P., & Bechara, A. (2010). The impact of prior risk experiences on subsequent risky decision-making: the role of the insula. *NeuroImage*, *50*, 709–716. https://doi.org/10.1016/j.neuroimage.2009.12.097

Yael, D., Zeef, D. H., Sand, D., Moran, A., Katz, D. B., Cohen, D., . . . Bar-Gad, I. (2013). Haloperidol-induced changes in neuronal activity in the striatum of the freely moving rat. *Frontiers in Systems Neuroscience*, *7*, 110. https://doi.org/10.3389/fnsys.2013.00110

Yang, C. R., & Seamans, J. K. (1996). Dopamine D1 receptor actions in layers V-VI rat prefrontal cortex neurons in vitro: modulation of dendritic-somatic signal integration. *The Journal of Neuroscience*, *16*, 1922–1935.

Yavich, L., Forsberg, M. M., Karayiorgou, M., Gogos, J. A., & Männistö, P. T. (2007). Site-specific role of catechol-O-methyltransferase in dopamine overflow within prefrontal cortex and dorsal striatum. *The Journal of Neuroscience*, *27*, 10196–10209. https://doi.org/10.1523/JNEUROSCI.0665-07.2007

Yoest, K. E., Quigley, J. A., & Becker, J. B. (2018). Rapid effects of ovarian hormones in dorsal striatum and nucleus accumbens. *Hormones and Behavior.* Advance online publication. https://doi.org/10.1016/j.yhbeh.2018.04.002

Yoshida, W., & Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. *Neuron*, *50*, 781–789. https://doi.org/10.1016/j.neuron.2006.05.006

Youngren, K. (1999). Clozapine Preferentially Increases Dopamine Release in the Rhesus Monkey Prefrontal Cortex Compared with the Caudate Nucleus. *Neuropsychopharmacology*, *20*, 403–412. https://doi.org/10.1016/S0893-133X(98)00082-7

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., & Kahana, M. J. (2009). Human substantia nigra neurons encode unexpected financial rewards. *Science*, *323*, 1496–1499. https://doi.org/10.1126/science.1167342

Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *ELife*, *6.* https://doi.org/10.7554/eLife.27430

Zappia, M., & Nicoletti, A. (2010). The role of the long-duration response to levodopa in Parkinson's disease. *Journal of Neurology*, *257*, S284-7. https://doi.org/10.1007/s00415-010-5731-0

Zeeb, F. D., Floresco, S. B., & Winstanley, C. A. (2010). Contributions of the orbitofrontal cortex to impulsive choice: interactions with basal levels of impulsivity, dopamine signalling, and reward-related cues. *Psychopharmacology*, *211*, 87–98. https://doi.org/10.1007/s00213-010-1871-2

Zetterström, T., Sharp, T., Marsden, C. A., & Ungerstedt, U. (1983). In Vivo Measurement of Dopamine and Its Metabolites by Intracerebral Dialysis: Changes After d-Amphetamine. *Journal of Neurochemistry*, *41*, 1769–1773. https://doi.org/10.1111/j.1471-4159.1983.tb00893.x

Zhang, L., Doyon, W. M., Clark, J. J., Phillips, P. E. M., & Dani, J. A. (2009). Controls of tonic and phasic dopamine transmission in the dorsal and ventral striatum. *Molecular Pharmacology*, *76*, 396–404. https://doi.org/10.1124/mol.109.056317

Zhang, S., & Yu, A. J. (2013). Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 2607–2615). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/5180-forgetful-bayes-and-myopic-planning-human-learning-and-decision-making-in-a-bandit-setting.pdf

Zhang, T., Di Mou, Wang, C., Tan, F., Jiang, Y., Lijun, Z., & Li, H. (2015). Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *International Journal of Psychophysiology*, *96*, 155–161. https://doi.org/10.1016/j.ijpsycho.2015.04.010

Zhuang, X., Oosting, R. S., Jones, S. R., Gainetdinov, R. R., Miller, G. W., Caron, M. G., & Hen, R. (2001). Hyperactivity and impaired response habituation in hyperdopaminergic mice. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 1982–1987. https://doi.org/10.1073/pnas.98.4.1982

Zimmermann, P., & Fimm, B. (2012). *Testbatterie zur Aufmerksamkeitsprüfung (TAP) – Version 2.3*. Herzogenrath, Germany: Psytest.

Zucchini, W. (2000). An Introduction to Model Selection. *Journal of Mathematical Psychology*, *44*, 41–61. https://doi.org/10.1006/jmps.1999.1276

Zweifel, L. S., Parker, J. G., Lobb, C. J., Rainwater, A., Wall, V. Z., Fadok, J. P., . . . Palmiter, R. D. (2009). Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *106*

14 Acknowledgment

First of all, I would like to express my sincere gratitude to my advisor Prof. Dr. Jan Peters for his support and guidance during these last years, for everything I learned from him, and for giving me the opportunity to work on this exciting and multifaceted project, which matches my scientific interests in so many respects.

Moreover, I would like to thank Prof. Dr. Steffen Moritz and PD Dr. Kirsten Hötting for agreeing to be in the dissertation committee, as well as the members of my thesis committee, Prof. Dr. Jürgen Gallinat and PD Dr. Stefanie Brassen.

Furthermore, I am very thankful to our study physician Dr. Florian Ganzer for his shared medical knowledge and significant contribution to putting this pharmacological fMRI experiment into practice. With respect to the fMRI scanning, a big thank you also to our MR physicist Dr. Jürgen Finsterbusch and to the skilled and sympathetic "trio" at the Trio (MR scanner), namely Katrin Bergholz, Kathrin Wendt, and Waldemar Schwarz. Many thanks also to all the participants who volunteered to take part in this study.

I am also very grateful to a large group of colleagues and friends, inside and outside the institute, for making the work on this PhD project a very inspiring and companionable experience. In particular, I would like to express my gratitude to Antonius Wiehler, who continuously helped me to tackle methodological challenges of this project and from whom I learned a lot about cognitive modeling and fMRI. I would also like to thank Dominique Goltz, Uli Bromberg, Julia Rihm, and Heidrun Schultz for their support in getting the project started and for many fruitful scientific discussions. Finally, a big thank you to Dr. Tobias Sommer and the members of his research group, especially Janine Bayer and Mareike Clos, for always welcoming me to their most enjoyable and interesting group meetings, and also for supporting me in many aspects of this thesis.

Last but not least, a very heartfelt thank you to my partner and family for their support, encouragement, and shared interest in my work.

15 Appendix

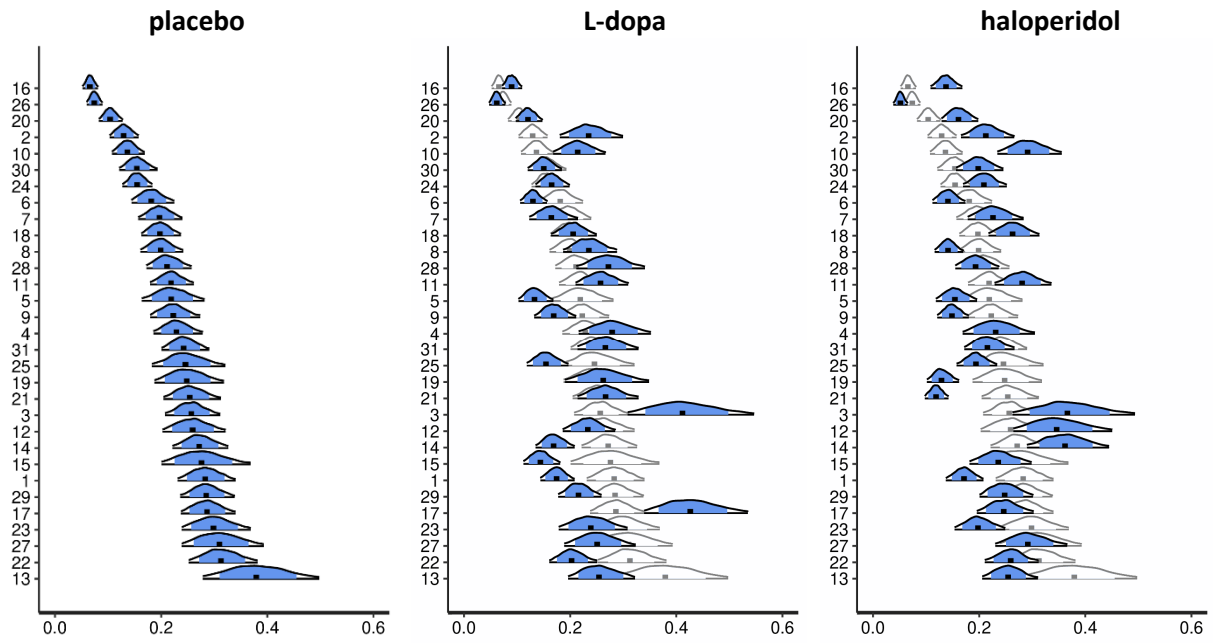


Figure A1. Subject-level parameter estimates for the softmax parameter (β). Shown are posterior distributions of the subject-level β parameter of the Bayes-SM+EP model, separately for each drug condition. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (blue area), and the 95 % central interval (black contours). For the L-dopa and haloperidol condition, posterior distributions (in blue) are overlaid on the posterior distributions of the placebo condition (in white) for better comparison.

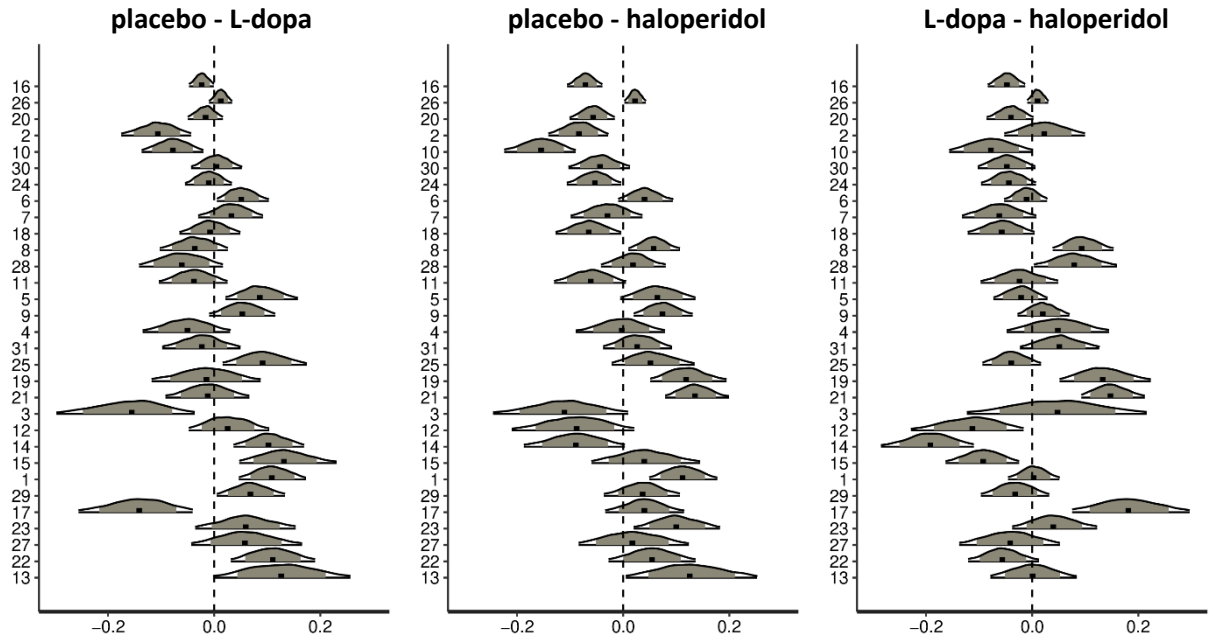


Figure A2. Drug effects on the softmax parameter (β) on the subject level. Shown are posterior drug differences of the subject-level β parameter of the Bayes-SM+EP model. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (grey area), and the 95 % central interval (black contours).

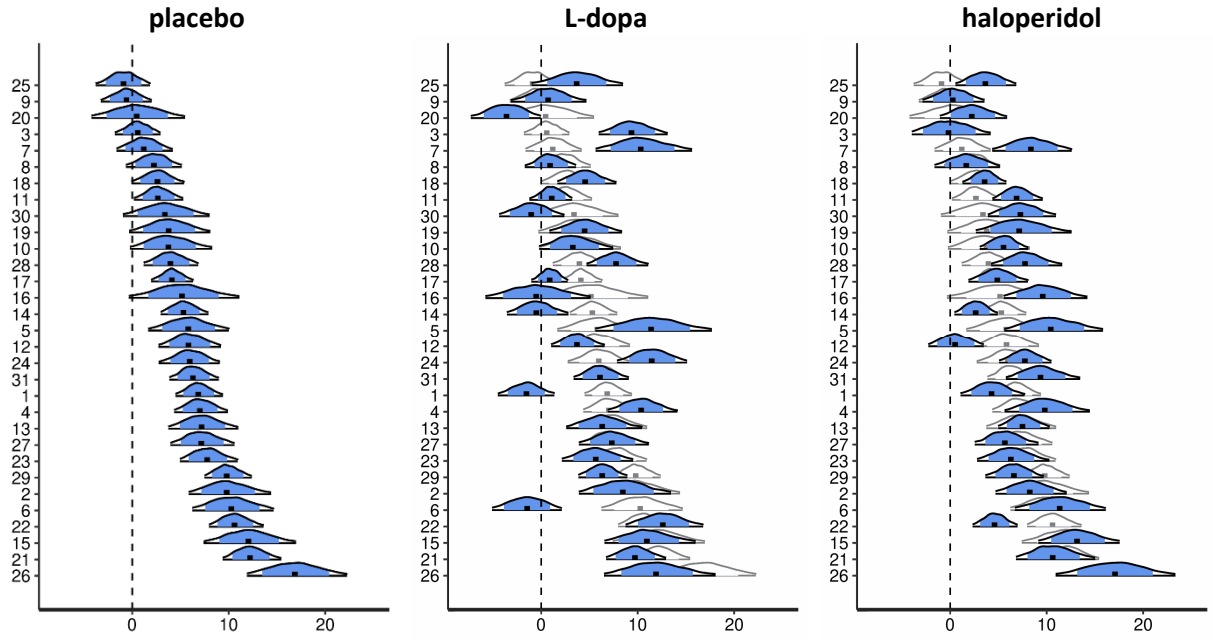


Figure A3. Subject-level parameter estimates for the perseveration bonus parameter (ρ). Shown are posterior distributions of the subject-level ρ parameter of the Bayes-SM+EP model, separately for each drug condition. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (blue area), and the 95 % central interval (black contours). For the L-dopa and haloperidol condition, posterior distributions (in blue) are overlaid on the posterior distributions of the placebo condition (in white) for better comparison.

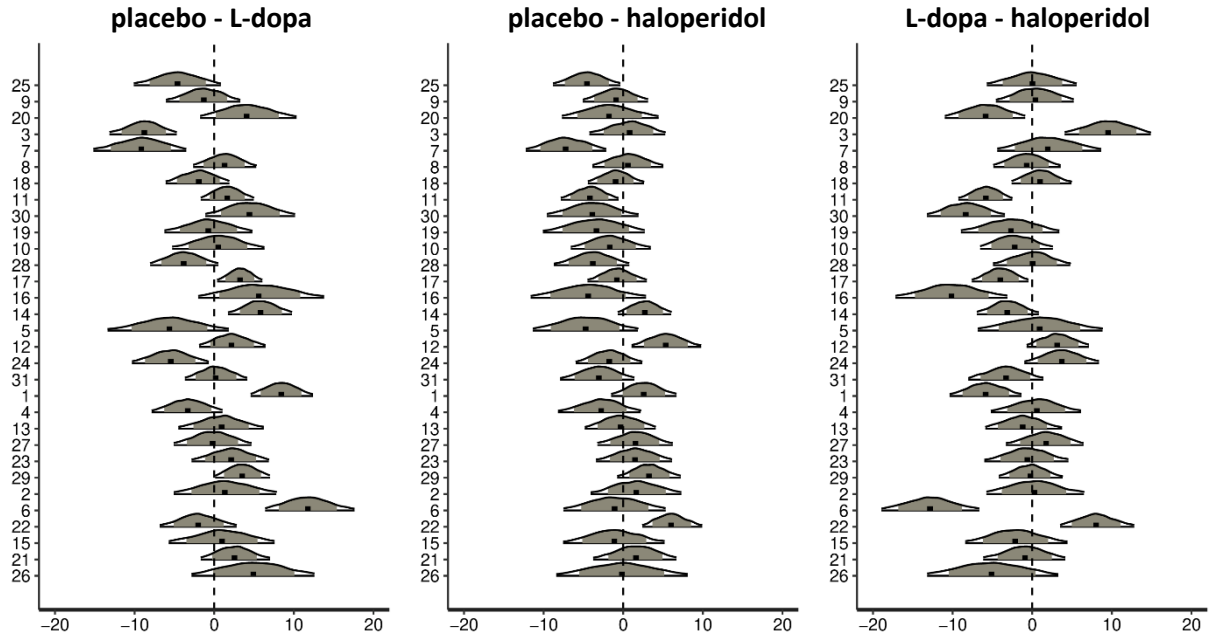


Figure A4. Drug effects on the perseveration bonus parameter (ρ) on the subject level. Shown are posterior drug differences of the subject-level ρ parameter of the Bayes-SM+EP model. For each posterior distribution, the plot shows the median (black dot), the 80 % central interval (grey area), and the 95 % central interval (black contours).

Table A1. Posterior medians of the random walk parameters in the Bayes-SM+EP model.

	$\hat{\lambda}$	$\hat{\vartheta}$	$\hat{\sigma}_d$	$\hat{\mu}_1^{pre}$	$\hat{\sigma}_1^{pre}$
pilot study (n=16)	0.93	45.99	6.62	82.72	3.61
placebo (n=31)	0.93	49.14	7.04	74.02	6.63
L-dopa (n=31)	0.93	51.81	6.79	70.80	9.26
haloperidol (n=31)	0.91	48.99	6.35	72.05	9.60

Note. Each random walk parameter was estimated once over all subjects in the respective sample. $\hat{\lambda}$: decay parameter; $\hat{\vartheta}$: decay center; $\hat{\sigma}_d$: diffusion standard deviation; $\hat{\mu}_1^{pre}$: initial prior mean of the expected reward for all bandits; $\hat{\sigma}_1^{pre}$: initial prior standard deviation of the expected reward for all bandits.

Table A2. Comparison of control variables (first set) between drug conditions.

control variable (n = number of subjects)	**placebo** mean (*SD*)	**L-dopa** mean (*SD*)	**haloperidol** mean (*SD*)	**ANOVA**
sEBR (n=30)	81.3 (94.6)	77.7 (76.0)	88.2 (77.0)	$F_{2,58} = 1.31$; $p = .278$
Digit Span Task				
forward: total score (n=31)	8.8 (2.2)	9.2 (1.8)	8.5 (1.9)	$F_{2,60} = 3.02$; $p = .056$
backward: total score (n=29)	8.6 (2.4)	7.9 (2.6)	7.9 (2.4)	$F_{2,56} = 3.09$; $p = .054$
TAP subtest Alertness (n=31)				
overall: RT, median	217.4 (19.1)	221.8 (21.5)	222.6 (31.5)	$F_{2,60} = 1.37$; $p = .263$
overall: RT, *SD*	32.9 (14.0)	33.7 (17.3)	37.9 (23.2)	$F_{2,60} = 1.73$; $p = .187$
overall: errors (misses)	2.10 (2.34)	2.19 (2.64)	2.23 (2.06)	$F_{2,60} = 0.06$; $p = .946$
intrinsic alertness: RT, median	218.3 (21.5)	221.4 (18.5)	223.3 (30.2)	$F_{2,60} = 1.22$; $p = .303$
intrinsic alertness: RT, *SD*	30.7 (13.3)	32.0 (14.2)	37.1 (23.3)	$F_{2,60} = 2.43$; $p = .097$
phasic arousal: RT, median	217.7 (18.5)	223.6 (33.1)	222.5 (34.9)	$F_{2,60} = 0.92$; $p = .406$
phasic arousal: RT, *SD*	34.1 (16.1)	31.8 (16.4)	36.0 (22.6)	$F_{2,60} = 0.95$; $p = .391$
index of phasic arousal	0.001 (0.04)	-0.005 (0.10)	0.006 (0.08)	$F_{2,60} = 0.20$; $p = .822$
TAP subtest Go/NoGo (n=31)				
RT, median	363.7 (55.8)	357.6 (45.3)	359.4 (53.9)	$F_{2,60} = 0.31$; $p = .732$
RT, *SD*	74.5 (24.8)	75.1 (18.0)	77.7 (29.1)	$F_{2,60} = 0.32$; $p = .726$
false alarms	1.39 (1.56)	1.68 (1.72)	1.77 (1.54)	$F_{2,60} = 1.39$; $p = .256$
misses	0.065 (0.25)	0.065 (0.36)	0.065 (0.25)	$F_{2,60} = 0$; $p = 1$
TAP subtest Flexibility (n=30)				
RT, median	530.9 (111.5)	546.5 (136.1)	527.1 (108.3)	$F_{2,58} = 0.97$; $p = .385$
RT, *SD*	138.7 (58.8)	139.4 (61.8)	135.2 (59.9)	$F_{2,58} = 0.15$; $p = .862$
errors	2.93 (2.61)	3.43 (3.80)	3.30 (2.90)	$F_{2,58} = 0.38$; $p = .688$

Note. The last column shows the result of the univariate repeated measures ANOVA with the factor drug for each control variable. RT: reaction time (in ms); sEBR: spontaneous eye blink rate; TAP: Tests of Attentional Performance (Zimmermann & Fimm, 2012).

Table A3. Comparison of control variables (second set) between drug conditions.

control variable (n = number of subjects)	**placebo** mean (*SD*)	**L-dopa** mean (*SD*)	**haloperidol** mean (*SD*)	**ANOVA**
side effects sum score (n=31)				
t_1 (1.0 h)	0.13 (1.26)	0.10 (2.47)	-0.03 (1.11)	$F_{2,60} = 0.08$; $p = .920$
t_2 (2.5 h)	-0.06 (1.36)	0.13 (1.93)	-0.23 (1.09)	$F_{2,60} = 0.59$; $p = .559$
t_3 (4.0 h)	-0.03 (1.35)	0.71 (1.97)	-0.06 (1.59)	$F_{2,60} = 2.77$; $p = .071$
vital parameters (n=31)				
pulse				
t_1 (1.0 h)	-7.65 (10.50)	-8.94 (9.18)	-7.29 (6.88)	$F_{2,60} = 0.31$; $p = .735$
t_2 (2.5 h)	-14.87 (10.11)	-15.32 (9.20)	-14.68 (11.56)	$F_{2,60} = 0.04$; $p = .963$
t_3 (4.0 h)	-14.61 (9.76)	-11.68 (11.15)	-12.94 (13.95)	$F_{2,60} = 0.66$; $p = .522$
diastolic blood pressure				
t_1 (1.0 h)	3.26 (7.94)	1.29 (7.39)	1.00 (6.86)	$F_{2,60} = 0.91$; $p = .407$
t_2 (2.5 h)	6.06 (8.16)	6.84 (7.51)	3.48 (6.81)	$F_{2,60} = 1.93$; $p = .155$
t_3 (4.0 h)	4.68 (7.84)	2.10 (8.55)	3.61 (7.63)	$F_{2,60} = 1.20$; $p = .309$
systolic blood pressure				
t_1 (1.0 h)	-1.42 (13.57)	-0.06 (10.91)	-0.45 (10.81)	$F_{2,60} = 0.13$; $p = .876$
t_2 (2.5 h)	2.45 (11.49)	3.16 (8.06)	0.52 (8.49)	$F_{2,60} = 0.82$; $p = .445$
t_3 (4.0 h)	1.87 (13.55)	3.81 (11.59)	3.65 (8.89)	$F_{2,60} = 0.33$; $p = .721$
mood ratings by VAS (n=31)				
alertness				
t_1 (2.5 h)	-0.08 (2.24)	-0.06 (3.03)	-0.17 (2.93)	$F_{2,60} = 0.01$; $p = .986$
t_2 (4.0 h)	2.19 (4.16)	2.50 (4.27)	2.22 (4.54)	$F_{2,60} = 0.28$; $p = .758$
contentedness				
t_1 (2.5 h)	0.09 (1.39)	0.02 (1.43)	0.07 (1.25)	$F_{2,60} = 0.13$; $p = .882$
t_2 (4.0 h)	0.24 (2.37)	0.37 (1.60)	0.20 (1.71)	$F_{2,60} = 0.08$; $p = .923$
calmness				
t_1 (2.5 h)	0.14 (0.96)	-0.18 (1.10)	0.11 (0.73)	$F_{2,60} = 1.09$; $p = .344$
t_2 (4.0 h)	-0.38 (1.48)	-0.56 (1.32)	0.29 (1.19)	$F_{2,60} = 4.46$; $p = .016$
mood ratings by SAM (n=31)				
pleasure				
t_1 (2.5 h)	0.10 (0.65)	-0.06 (0.93)	0.16 (0.78)	$F_{2,60} = 0.79$; $p = .458$
t_2 (4.0 h)	0.03 (0.84)	0.33 (1.09)	0.19 (0.83)	$F_{2,60} = 1.40$; $p = .254$
arousal				
t_1 (2.5 h)	-0.26 (0.89)	-0.23 (0.99)	-0.19 (0.75)	$F_{2,60} = 0.04$; $p = .956$
t_2 (4.0 h)	0.26 (0.77)	0.10 (1.12)	-0.06 (0.93)	$F_{2,60} = 1.55$; $p = .220$
dominance				
t_1 (2.5 h)	0.06 (0.44)	-0.03 (0.55)	-0.03 (0.60)	$F_{2,60} = 0.47$; $p = .630$
t_2 (4.0 h)	0.03 (0.71)	-0.03 (0.81)	-0.23 (0.72)	$F_{2,60} = 1.03$; $p = .364$

Note. All variables are reported as difference measures relative to a baseline (t_0) assessed directly before drug administration. The last column shows the result of the univariate repeated measures ANOVA with the factor drug for each control variable. VAS: Visual Analogue Scale (Bond & Lader, 1974); SAM: Self-Assessment Manikin (Lang, 1980).

Table A4. Correlations between all model-based fMRI regressors.

	explore (overall)	explore (directed)	explore (random)	expected value	uncer-tainty	overall uncertainty	prediction error
explore (overall)[a]	-	.55	.71	-.63	.78	-.25	.02
explore (directed)[b]	.55	-	-.10	-.40	.59	-.19	.00
explore (random)[b]	.71	-.10	-	-.43	.53	-.11	.03
expected value[c]	-.63	-.40	-.43	-	-.62	.29	-.05
uncer-tainty[c]	.78	.59	.53	-.62	-	-.19	.06
overall uncertainty[d]	-.25	-.19	-.11	.29	-.19	-	.13
prediction error[e]	.02	.00	.03	-.05	.06	.13	-

Note. Reported are Pearson correlation coefficients across all trials, subjects, and drug conditions. Correlation coefficients for *exploit* (first GLM) are the same as reported for *explore (overall)*, but with inverse sign.

[a] first (main) GLM; [b] second GLM; [c] third GLM; [d] fourth GLM; [e] included in all GLMs.

Table A5. Regions used for small volume correction.

region of small volume correction	**peak voxel (mm)**			**reference for peak voxel**
	x	**y**	**z**	
right frontopolar cortex (rFPC)	27	57	6	Daw et al. (2006)
left frontopolar cortex (lFPC)	-27	48	4	Daw et al. (2006)
right intraparietal sulcus (rIPS)	39	-36	42	Daw et al. (2006)
left intraparietal sulcus (lIPS)	-29	-33	45	Daw et al. (2006)
right anterior insula (rAI)	32	22	-8	Blanchard & Gershman (2018)
left anterior insula (lAI)	−30	16	-8	Blanchard & Gershman (2018)
dorsal anterior cingulate cortex (dACC)	8	16	46	Blanchard & Gershman (2018)

Note. Each small volume correction used a 10-mm-radius sphere around the listed peak voxel MNI coordinates, which mark brain regions that have previously been associated with exploratory choices.

Table A6. Brain regions showing higher activity for exploratory than exploitative choices (first GLM).

Region	MNI coordinates			peak	cluster
	x	y	z	z-value	extent (k)
R/L intraparietal sulcus, R/L precuneus, R/L postcentral gyrus, L precentral gyrus	-48	-33	52	10.45	15606
R precentral gyrus	26	-8	50	9.32	2297
R/L supplementary motor cortex, R/L dorsal anterior cingulate cortex	8	12	45	8.47	2552
R cerebellum, R fusiform gyrus	18	-51	-22	8.09	2574
R middle frontal gyrus	39	34	28	7.56	1291
R cerebellum	24	-57	-54	7.35	128
L precentral gyrus	-51	0	34	7.31	430
L cerebellum, L fusiform gyrus	-40	-54	-32	7.28	1419
L thalamus	-10	-20	6	6.96	556
R/L calcarine cortex	-8	-74	14	6.90	1222
R anterior insula	36	20	3	6.87	511
L anterior insula	-36	15	3	6.69	557
R precentral gyrus	51	8	24	6.49	434
R thalamus	10	-18	8	6.32	331
R cerebellum	30	-44	-48	6.24	28
L middle frontal gyrus	-42	27	27	6.07	97
R cerebellum	14	-62	-45	5.88	61
R pallidum	15	6	-4	5.83	25
R calcarine cortex	9	-94	6	5.74	104
vermis	3	-75	-34	5.70	52
R supramarginal gyrus	51	-42	28	5.69	46
L middle frontal gyrus	-30	46	15	5.67	47
L pallidum	-10	6	-4	5.64	51
R anterior orbital gyrus	24	54	-9	5.60	33
L posterior cingulate cortex	-3	-32	26	5.51	21
L caudate nucleus	-16	-14	18	5.33	28
R caudate nucleus	12	-8	16	5.24	16
L lingual gyrus	-16	-84	-12	5.21	10
R anterior cingulate cortex	10	27	21	5.13	10

Note. Thresholded at $p < .05$, FWE-corrected for whole-brain volume, with $k \geq 10$ voxels. L: left; R: right.

Table A7. Brain regions showing higher activity for exploitative than exploratory choices (first GLM).

Region	MNI coordinates			peak	cluster
	x	y	z	z-value	extent (k)
L angular gyrus	-42	-74	34	8.04	2530
L posterior cingulate cortex, L precuneus	-6	-52	15	7.40	1087
R angular gyrus	52	-68	28	7.02	185
R postcentral gyrus	33	-26	54	6.80	503
R cerebellum	27	-78	-38	6.28	452
R rostral anterior cingulate cortex	4	18	-14	5.90	125
L superior temporal gyrus	-62	-36	3	5.89	70
L lateral orbital gyrus	-38	34	-14	5.81	102
R central operculum	45	-14	20	5.73	83
L middle temporal gyrus	-62	-4	-22	5.67	193
R/L medial frontal cortex	-2	40	-10	5.67	233
L superior frontal gyrus	-10	54	30	5.54	20
L superior frontal gyrus	-10	51	36	5.45	10
L middle temporal gyrus	-60	-51	-2	5.38	61
R superior temporal gyrus	52	-12	-9	5.35	25
R middle temporal gyrus	62	4	-21	5.30	10
L rostral anterior cingulate cortex	-6	46	4	5.17	13
L inferior frontal gyrus	-50	27	2	5.16	20

Note. Thresholded at $p < .05$, FWE-corrected for whole-brain volume, with $k \geq 10$ voxels. L: left; R: right.

Table A8. Brain regions in which activity was significantly correlated with the overall uncertainty (fourth GLM), shown for the placebo condition and for pairwise drug comparisons.

Region	MNI coordinates			peak	cluster
	x	y	z	z-value	extent (k)
placebo					
L posterior insula	-34	-20	8	4.63	198
R supplementary motor cortex	8	10	52	3.98	92
R/L dorsal anterior cingulate cortex, L supplementary motor cortex	-3	21	39	3.96	176
R anterior insula	42	15	-6	3.46	38
R thalamus	8	-10	2	3.41	18
placebo > L-dopa					
L posterior insula	-34	-20	8	5.05 [a]	82
L anterior insula, L frontal operculum	-38	6	14	4.88	222
L opercular part of inferior frontal gyrus	-42	9	26	4.01	80
L precentral gyrus	-54	3	12	3.47	23
R dorsal anterior cingulate cortex	4	14	28	3.41	32
R precentral gyrus	39	-9	44	3.39	16
L dorsal anterior cingulate cortex	-2	36	33	3.32	17
L-dopa > placebo					
no suprathreshold activation					
placebo > haloperidol					
R/L thalamus	2	-10	-2	4.52	115
L posterior orbital gyrus	-24	30	-14	4.17	63
L cerebellum	-33	-48	-45	4.14	117
L parahippocampal gyrus	-22	-30	-21	3.91	21
R anterior insula, R frontal operculum	39	15	-4	3.88	125
L anterior insula	-30	21	6	3.81	27
L posterior insula	-34	-22	8	3.78	31
L central operculum	-54	-20	18	3.73	95
L hippocampus	-27	-38	-9	3.72	36
R cerebellum	28	-42	-46	3.67	66
R lingual gyrus	21	-40	-15	3.64	24
L precentral gyrus	-30	-18	44	3.64	14
L cerebellum	-30	-64	-56	3.50	23
R precentral gyrus	15	-26	45	3.49	17
R parietal operculum	48	-28	27	3.48	43
L frontal operculum	-50	14	-4	3.46	41
R middle temporal gyrus	56	-54	-4	3.44	16

R superior temporal gyrus	62	-38	14	3.43	11
L cerebellum	-40	-62	-51	3.41	13
L lingual gyrus	-9	-64	-3	3.40	11
L cerebellum	-15	-69	-50	3.37	20
L cerebellum	-28	-62	-30	3.37	14
R precentral gyrus	46	-10	44	3.36	13
R temporal pole	54	12	-8	3.35	16
R precentral gyrus	46	6	22	3.31	10
L cerebellum	-34	-64	-46	3.29	17
haloperidol > placebo					
no suprathreshold activation					
L-dopa > haloperidol					
no suprathreshold activation					
haloperidol > L-dopa					
no suprathreshold activation					

Note. Thresholded at $p < .001$, uncorrected, with $k \geq 10$ voxels. L: left; R: right.

[a] $p = .031$, FWE-corrected for whole-brain volume.

16 Curriculum Vitae

Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

17 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: _____